Measuring Behavior 2014

9th International Conference on Methods and Techniques in Behavioral Research

27-29 August 2014, Wageningen, The Netherlands

Proceedings

Proceedings of

Measuring Behavior 2014

9th International Conference on Methods and Techniques in Behavioral Research

Wageningen, The Netherlands
August 27-29, 2014

Editors
Andrew Spink
Egon L. van den Broek
Leanne Loijens
Marta Woloszynowska-Fraser
Lucas Noldus

Noldus
Information Technology

2014

Bibliographical reference

Proceedings of Measuring Behavior 2014: 9th International Conference on Methods and Techniques in Behavioral Research (Wageningen, The Netherlands, August 27-29, 2014)

Edited by A.J. Spink, E.L. van den Broek, L.W.S. Loijens, M. Woloszynowska-Fraser and L.P.J.J. Noldus

Copyright

P.O. Box 268

6700 AG Wageningen, The Netherlands

http://www.noldus.com/

Table of contents

Keynotes

Oral presentations

General Sessions

Posters

Welcome to *Measuring Behavior 2014*

We are delighted to welcome you all to the 9th *Measuring Behavior* conference, which takes place in the city of Wageningen. Wageningen is one of the oldest cities in The Netherlands, having just celebrated 750 years of city rights last year. The University was founded in 1876, building on the local Agricultural College which already existed. It is new a world-center for biological research, with several groups doing outstanding research in both animal and human behavior as well as a variety of commercial spin-off companies (including the long-term sponsor of *Measuring Behavior*, Noldus Information Technology BV).

Building on the format that has emerged from previous meetings, we have a fascinating program of presentations about a wide variety of methodological aspects of the behavioral sciences. In addition to purely scientific presentations, we also have a good number of presentations in the Demonstration Showcase in which academies and companies show off their latest prototypes.

The scientific program contains contributions focusing on purely scientific aspects (issues of replica-bility, dynamic aspects of behavior) and applied research (animal welfare), human behavior (eye trackers in consumer research) and animal (rodent behavior), technical sessions (video tracking of social animals and recognition of human behaviors from video), sessions presenting the latest tech-nology (3D simulators) and topics that are of relevance to everyone (eating behavior of people). The above list just scratches the surface of what promises to be a very diverse and interesting three days.

We hope this program caters for many of your interests and we look forward to seeing and hearing your contributions and trust it will become a productive, exciting and memorable conference.

Andrew Spink (Noldus Information Technology, The Netherlands)

Gemot Riedel (University of Aberdeen, United Kingdom)

Egon L. van den Broek (Utrecht University, The Netherlands)

Maurizio Mauri (University Institute for Modern Languages, ltaly)

Conference Chairs

The *Measuring Behavior* Conferences

Measuring Behavior is a unique conference about methods and techniques in behavioral research. While most conferences focus on a specific domain, *Measuring Behavior* creates bridges between disciplines by bringing together people who may otherwise be unlikely to meet each other. At a *Measuring Behavior* meeting, you find yourself among ethologists, behavioral ecologists, neuroscientists, experimental psychologists, human factors researchers, movement scientists, robotics engineers, software designers, electronic engineers, human-computer interaction specialists… to name but a few. Experience tells us that the focus on methodological and technical themes can lead to a very productive cross-fertilization between research fields. Crossing the boundaries between disciplines and species (from astronauts to zebras) can be extremely inspiring.

Measuring Behavior started in 1996 as a workshop in the framework of a European research project "Automatic Recording and Analysis of Behavior", aimed at sharing the results of our project with colleagues from abroad. Organized by Noldus Information Technology and hosted by Utrecht University, *Measuring Behavior '96* attracted over 150 participants. From that modest beginning, the conference has grown to a significant international event with several hundred delegates from thirty plus countries. We have also grown in terms of the scientific quality of the conference, with selection of papers now being determined by a process of independent peer-review by many hundreds of reviewers. The scientific program committee is very grateful for all that work that many of you reading this have contributed towards.

Noldus Information Technology serves as conference organizer and main sponsor. For a small company like ours, the conference is a major investment. We gladly do this, because we believe that the focused attention on behavior research methods and techniques will eventually lead to a higher demand for our solutions. To prevent commercial bias, however, the scientific program is put together under auspices of an independent Scientific Program Committee, consisting of international experts from a broad variety of disciplines (see the Scientific Program Committee in the Acknowledgments section) and many different research groups and companies have contributed and participated in this series of conferences.

Over the years, the conference has been hosted by a variety of universities:

Year	City	Conference chair
1996	Utrecht	Berry Spruijt
1998	Groningen	Jaap Koolhaas
2000	Nijmegen	Alexander Cools
2002	Amsterdam	Gerrit van der Veer
2005	Wageningen	Louise Vet
2008	Maastricht	Harry Steinbusch
2010	Eindhoven	Boris de Ruyter
2012	Utrecht	Remco Veltkamp, Gernot Riedel
2014	Wageningen	Gemot Riedel, Egon L. van den Broek, and Maurizio Mauri

In the scientific program, which is well balanced between human and animal research, you can find a variety of formats for presentation, interaction and exchange of information. In the past years we have seen that the special sessions (with speakers invited by session chairs) have become more prominent, and also the Demonstration Showcase has become more popular.

Measuring Behavior is a scientific conference, so special attention is paid to publication of the work presented at the meeting. The format of papers in the Proceedings is always a matter of debate, due to the different conventions of the many disciplines represented at the conference. After trying a variety of formats over the years, we have settled on extended abstracts (1-2 pages for posters and 2-5 pages for oral presentations), and that seems to be the best compromise between a text with enough content for both lasting value and being possible to review and short enough so that the effort for the writers is not excessive. An important feature of the conference Proceedings is that they are all available as open access from http://www.measuringbehavior.org.

Now you find yourself at the 9th *Measuring Behavior* conference. The organizers have done their best to prepare an optimal mix of scientific, technical, and social ingredients. We hope that you will find *Measuring Behavior 2014* a rewarding and stimulating experience and wish you a pleasant stay in Utrecht.

Keynotes

Strategies in spatial learning:

New in-depth psychological analysis using old tools

Gernot Riedel, Marta Urszula Woloszynowska-Fraser, and Barry Crouch

School of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK.
g.riedel@abdn.ac.uk

Abstract

The overarching interest of our current research focusses on the detection of global changes in neuronal activity measured as electroencephalogram (EEG). In disease models, we now have evidence that spatial learning is severely compromised and a more in-depth analysis of which cognitive strategy is implemented has become a pre-requisite for further detailed understanding of how the global network activity can bring about behaviour.

Here, we describe several attempts to devlop an automated form of strategy classification for data taken from our substantial archive of water maze experiments (>1500 tracks). We first implemented a manual visual analysis and explored within-experimenter and between-experimenter reliability. While the former achieved a level of accuracy of >80%, the latter decayed to marginally more than 60%. Automated classification using Fuzzy Logic proved less reliable, but classification based on Mahalanobis Distance (outlined here) achieved overall accuracies matching the intra-scorer values of 80% correct.

Our novel algorithmic approach to automated analysis of mouse visuospatial search strategy uses a recently developed Matlab script. This method operates through an analysis of only the raw x-y co-ordinates of an animal's track and can therefore be utilised independent of which video tracking package is used to record behaviour. The method operates in two stages: first a number of track parameters such as area coverage, path efficiency and thigmotaxis are automatically calculated in order to represent each track as a string of numeric values (feature vector) before classification. The second stage computes probability of membership based on a cluster analysis while correcting for variance and covariance.

This approach can handle huge data sets and autoscore strategies executed by each individual throughout learning. It is admissable to water maze protocols, but also for other tasks using goal-related training protocols such as Barnes maze etc. Based on these strategy scores, we can now dissect EEGs and correlate behavioural failure with neuronal processing anomalies.

Background

Analysis of visuospatial search strategy is widely recognized as a powerful metric of rodent cognition in the open field water maze (WM). In the simplest terms a strategic analysis consists of classifying swim paths into one of a number of qualitatively similar 'pathforms'. There is little consistency in the literature as to the number of WM escape strategies presumed to exist or the conditions which characterize each strategy. However the more extensive and commonly cited classification schemes (Garthe, Behr, & Kempermann, 2009; Janus, 2004; Graziano, Petrosini, & Bartoletti, 2003; Lang, et al., 2003) can be broadly summarized as containing 3 main classes of search strategy, here termed Spatial, Procedural and Random. Commonly used sub-divisions of these classes include Direct, Focal and Directed searches within the spatial class and Chaining or scanning behaviour within the procedural class [Fig. 1]. So far, strategic analyses were predominantly carried out through a process of manual scoring. This almost by definition renders the process vulnerable to experimenter bias, human error and variability both between experimenters and within experimenters over time. These considerations alongside the value of strategic analysis are the driving force behind the development of automated pattern recognition systems which could greatly accelerate the process and remove human error (Wolfer, Madani, Valenti, & Lipp, 2001; Garthe & Kempermann, 2013; Graziano, Petrosini, & Bartoletti, 2003).

Methods

In order to provide a benchmark for evaluating classifier accuracy an examination of the efficacy of human scoring was conducted. 2 observers with similar levels of training were provided with an identical set of 200 unscored water-maze tracks. Each observer was required to score the data set twice with a 24 hour interval between scoring sessions. The disparity in scoring between each observer's day one and day 2 scoring was taken as indicative of within subject variance over time. The disparity of scoring between each both observers day 1 scores was taken as indicative of between observer variance when both observers are appropriately trained using the same guidelines.

Animals

All requisite data for construction and validation of the classifier was obtained with permission from a behavioural data archive. As such no new experiments were conducted specifically for the purpose of this study. WM Training data was obtained from the training of 77 mice in total of various wild type and disease model strains. Animal trials were recorded by an overhead camera connected to a PC running video tracking software (Anymaze, Ugo Basile). Raw time resolved x and y co-ordinates of the animals location over the duration of each trial were exported in individual comma separated value (.csv) files along with those of the target platform.

Strategy classes

WM tracks were classified using a 6 strategy classification scheme derived from an amalgam of those proposed by; Garthe, Behr, & Kempermann, 2009; Janus, 2004 and Graziano, Petrosini, & Bartoletti, 2003 [Fig. 1].

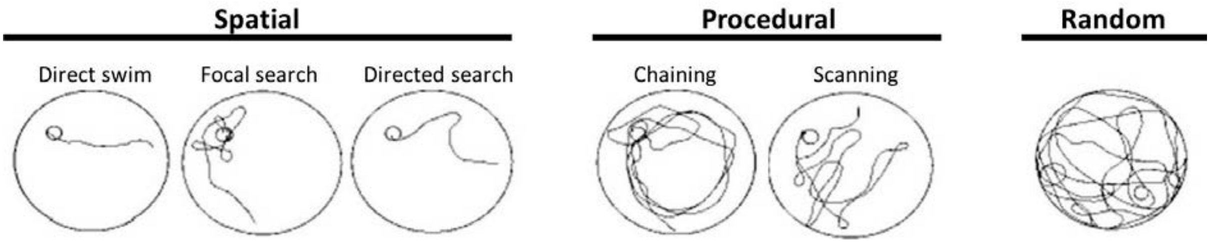


Figure 1. Water maze strategy classification scheme. Six strategy classifications were derived from an amalgam of those proposed by; Garthe, Behr, & Kempermann, 2009; Janus, 2004 and Graziano, Petrosini, & Bartoletti, 2003. Path forms are divided into 3 parent categories; the spatial category contains 3 pathforms termed **Direct**, **Focal** and **Directed.** The Procedural category contains **Chaining** and **Scanning** type pathforms. Pathforms with no dicernable use of strategy are termed **Random.**

Mahalanobis distance classification

For classification purposes each track is first represented mathematically as a series of 12 numeric parameters (Area coverage, path efficiency, mean distance to centre etc) termed the feature vector. Feature vectors are classified by a simple Mahalanobis distance (D^2) based classifier which computes D_M between a track feature vector (x) and the group mean vector (μ) of sample tracks of known class. This was accomplished by modelling each strategy class as a 12 dimensional Gaussian hyper-ellipsoid defied by the class μ and class covariance matrix (S) obtained from a manual scoring of >1000 WM tracks. D^2 carries the advantage of correcting for variance and covariance between variables.

$$D^2 = (x - \mu)^T S^{-1} (x - \mu)$$

Results

Human scoring accuracy

Diagnostics of human scoring illustrate that a relatively high degree of reproducibility is possible when a single observer is tasked with manual classification (Table 1A) as 80% of all tracks were assigned to the same class during both scoring sessions (chance = 16.7%). The majority of disparity is accounted for by within category substitutions (see Direct to Directed, Scanning to Chaining etc). When the day 1 scores of 2 different observers with similar training were compared (Table 1B), the total agreement was relatively poor (61%). While the number of outside category substitutions remains relatively low there is a much greater proportion of within category substitution. These results can be interpreted as an indication that ~40% of tracks are sufficiently ambiguous in class that two observers will produce a different classification and that ~20% of tracks are sufficiently ambiguous that the same observer will produce different scorings over multiple viewings. While differential scoring over multiple sessions is not possible with this method of automated scoring the within experimenter accuracy is indicative of training set quality. The within experimenter variability was therefore taken as the benchmark for validation of the classifier.

Table 1a. Confusion matrix of within experimenter variability.

A v A	Direct	Focal	Directed	Chaining	Scanning	Random
Direct	85.7	2.9	11.4	0.0	0.0	0.0
Focal	0.0	82.9	17.1	0.0	0.0	0.0
Directed	2.3	11.6	72.1	11.6	2.3	0.0
Chaining	0.0	4.5	0.0	90.9	4.5	0.0
Scanning	0.0	2.8	0.0	19.4	75.0	2.8
Random	0.0	0.0	0.0	0.0	20.7	79.3

Table 1b. Confusion matrix of between experimenter variability.

A v B	Direct	Focal	Directed	Chaining	Scanning	Random
Direct	60.0	0.0	40.0	0.0	0.0	0.0
Focal	0.0	60.0	25.7	0.0	14.3	0.0
Directed	2.3	14.0	39.5	2.3	41.9	0.0
Chaining	0.0	4.5	13.6	50.0	31.8	0.0
Scanning	0.0	11.1	0.0	11.1	69.4	8.3
Random	0.0	0.0	0.0	0.0	6.9	93.1

Notes. A) Rows - ground truth obtained from the day 1 scores of experimenter A, Columns - classifications obtained from day 2 scores of experimenter A. B) Rows - ground truth obtained from the day 1 scores of experimenter A, Columns - classifications obtained from day 1 scores of experimenter B. Matrix diagonals therefore represent % tracks correctly assigned accurate to 1dp.

Automated scoring accuracy

When a leave-one-out cross validation (LOOCV) diagnostic was used to determine expected classifier accuracy (Table 2), the results obtained were highly similar to that obtained with the diagnostic of within observer variability (Table 1A). This is somewhat to be expected given that this diagnostic is indicative of training data quality. Overall 84.6% of tracks were correctly assigned by the classifier. Focal, directed and random classes were detected with greater reproducibility (+0.1%, +18.6%, +8.7% respectively) than a single human observer while Direct, Chaining and Scanning classes were less consistently scored (-1.5%, -1.3% and -10% respectively). Scanning to chaining substitution was the most severely confused in both human and automated marking indicating the relatively poor resolution of these two categories may stem from a very high level of ambiguity between these two strategies and thus particularly high crossover in the training set. When compared with the between observer diagnostic (Table 1B), direct, focal, directed and chaining strategies were detected with a greatly enhanced reproducibility (+24.2%, +23%, 51.2% and 39.6% respectively) while scanning and random classes were less consistently scored (-4.4% and -5.1%). Overall these results indicate that the classifier performs at approximately human level accuracy with no drift in scoring over time.

Table 2. Confusion matrix of leave one out cross validation of classifier.

LOOCV	Direct	Focal	Directed	Chaining	Scanning	Random
Direct	84.2	1.3	14.6	0.0	0.0	0.0
Focal	0.7	83.0	14.3	2.0	0.0	0.0
Directed	1.0	6.8	90.7	1.5	0.0	0.0
Chaining	0.0	1.8	4.9	89.6	1.2	2.4
Scanning	0.0	0.0	0.0	27.5	65.0	7.5
Random	0.0	0.0	0.0	8.1	4.0	88.0

Notes. Rows - ground truth obtained from a manual scoring of the training data set, Columns - classifications obtained from a leave one out cross validation of the classifier operating on the same data set. Matrix diagonals therefore represent % tracks correctly assigned accurate to 1dp.

Discussion

The first issue which this project highlights is the inherent variability of human scoring even within a single experimenter over a short time. While this is the impetus behind the creation of automated systems it is also one of the primary challenges as without a definitive ground truth to calibrate against, it is difficult to asses the accuracy of a classifier. To elabourate our mahalanobis WM classifier produces an 80% match to human scoring, if the human scorer would produce only an 80% match on re-scoring the same data set, where then lies the truth of the matter? One could argue that by selecting only unambiguous 'archetypal' pathforms for a training set then this would partially resolve the issue and indeed using this method we have 'boosted' accuracy by 5-10%. But what then of the remainder, in real-world analysis the ambiguous tracks cannot be removed from analysis and must receive a classification. Alternatively one could discard the established classification schemes and could allow a cluster detection tool to create new strategy classses based on similarity of tracks without any *a priori* assumptions as to the number and nature of pathforms which may exist. However by essentially creating a new classification scheme for every data set would reduce the comparability between sussessive data sets and move toward diversification of analysis methods rather than toward consensus. Another hypothetical direction could be the collaborative creation of a large data set by first achieving a consensus on which classification scheme to use and then submitting tracks from multiple laboratories for co-operative classification. This data set could then be used by different laboratories to calibrate their individual classification tools thereby providing consistency in their operation. This approach however may be somewhat ideallistic and would be very difficult to orchastrate. One could also take the position that due to the ultra-high throughput analysis offered by automated classification (>1000 tracks per minute in this classifier), they carry immense value as a tool for pre-screening archives of historical experimental data to detect interesting trends prior to manual confirmation.

Regardless of ground truth this project demonstrates that human-level accuracy in automated scoring can be easily achieved with relatively low complexity systems provided that they account for variance and co-variance of parameters. Further we have shown that hese systems need not depend on any one video tracking software package provided that one can extract time resolved co-ordinates of a swim path. Similarly while the classifier presented here was written in the Matlab programming language for ease and speed of prototyping it is demonstrably reproducable in open source alternatives with relatively little modification. Implementing these systems therefore requires no monetary cost beyond that of a video tracking system. These methods have already been adapted for the Barnes maze (Barnes, 1979) and produce highly similar levels of accuracy.

References

[1] Barnes, C.A. (1979). Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *Journal of Comparative and Physiological Psychology,* ***93****(1),* 74-104.

[2] Gallagher, M., Burwell, R., Burchinal, M. (1993). Severity of spatial learning impairment in aging: development of a learning index for performance in the Morris water maze. *Behavioral Neuroscience,* ***107****(4),* 618-626.

[3] Garthe, A., Behr, J., Kempermann, G. (2009). Adult-generated hippocampal neurons allow the flexible use of spatially precise learning strategies." *PLoS One* ***4****(5)*, e5464.

[4] Garthe, A., Kempermann, G. (2013). An old test for new neurons: refining the Morris water maze to study the functional relevance of adult hippocampal neurogenesis. *Frontiers on Neuroscience,* ***7***, 63.

[5] Graziano, A., Petrosini, L., Bartoletti, A. (2003). Automatic recognition of explorative strategies in the Morris water maze. *Journal of Neuroscience Methods,* ***130****(1),* 33-44.

[6] Janus, C. (2004). Search strategies used by APP transgenic mice during navigation in the Morris water maze. *Learning & Memory,* ***11****(3),* 337-346.

[7] Lang, U.E., Lang, F., Richter, K., Vallon, V., Lipp, H.P., Schnermann, J., Wolfer, D.P. (2003). Emotional instability but intact spatial cognition in adenosine receptor 1 knock out mice. *Behavioural Brain Research* ***145****(1-2)*, 179-188.

[8] Wolfer, D.P., Madani, R., Valenti, P., Lipp, H.P. (2001). Extended analysis of path data from mutant mice using the public domain software Wintrack. *Physiology & Behavior,* ***73****(5),* 745-753.

Measuring behavior: In the moment and longer term

J. A. Healey

User Experience Innovations, Intel Labs, Santa Clara, CA, USA
jennifer.healey@intel.com

Introduction

Measuring behavior in living organisms is a difficult challenge. Considering that behavior is "the way in which an animal or person acts in response to a particular situation or stimulus," we can extrapolate that as the actor or the stimulus becomes more complex or subtle, the challenge of measuring the behavior likewise becomes more difficult. This paper calls out some of the difficulties associated with measuring human behavior and how these challenges might potentially be met by new technologies that enable measurements over longer time windows than might ever be studied in a lab and with more participants than any single study would ever have previously been able to traditionally recruit.

Measuring emotion

One of the most challenging types of behavior for scientists to measure in people is emotional behavior. To begin with, the ground truth is murky at best. The gold standard is the person's own self-report, but it is well understood that people are not always entirely in a position to judge their own emotional responses in the moment. People also have various levels of self-awareness; different cultural framing; different ways of wanting to be perceived and different sets of social rules when it comes to expressing the label or rating for their emotions. This may not just be a "rating" issue, but because emotion is at least partially subjective different the actual experience of emotion may different for different people. Consider that a tree rustling in the night may have terrified us as a child, but now the same stimulus does not instill terror because we frame the potential threat differently. Also, although we may have reported our true fear as a child to our mother, we might have told our older brother, "I'm not scared." Measuring emotion is also difficult in a complex organism like a person because emotion is just one component of the totality of a person's experience at any moment in time. Ultimately what you measure from a person is only the sum of responses to all of their experience.

Rating emotion

Anecdotally, I had one participant proclaim that rating emotion was impossible. He claimed that he had spent years in therapy trying to understand own his emotions and still felt that he was just beginning to understand his reactions. In my experience, this lack of awareness is common; people either do not really know how they are feeling or they express their experience using colloquial terms such as "I feel fine," or "I am happy to that is over." Emotion theorists have not extensively studies the emotion "fine" and the second response is closer to "relief" than the "enjoyment" defined by such emotion theorists as Ekman [1] but this subtlety is often not well understood by participants. Another difficulty for participants is understanding the taxonomy of time scales that theorists often associate with affect: emotion vs. mood vs. state or disposition. According to Ekman [1], emotions are short duration events, lasting several seconds, however, most people self-report their perception emotion events lasting closer to around half an hour on average[2]. There is also the question of the frequency at which emotional events occur. Are people always in some emotional state? If so, to what extent can they perceive it? According to a study by Oatley and Duncan perceived experience of emotion seems to be sparse for participants [3], with subject reporting approximately one memorable emotion per day, yet in many emotion monitoring experiments, the expectation is that subjects should be able to report an emotion whenever queried [4]. If queried during a time when they might not have self-perceived a particular emotional event normally, what is the "quality" of the result as an emotion? Might the response be overall dominated by disposition or mood? Is the simplest, best way to allow people to self-report an essentially highly complex state, especially while on the go in an ambulatory environment?

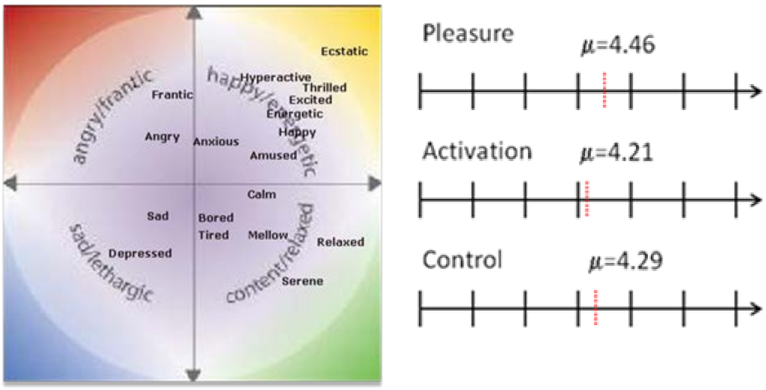


Figure 1. An illustration of the difference in means of emotion words from two experiments. The emotion words from the first experiment had a positive high energy bias. This was eliminated in a second experiment using only emotion axis descriptors.

Evidence	Acc
2 of 3 raters agreed	67%
In situ and end of day ratings agreed	79%
All three agreed	100%

Table 1. Machine learning algorithms perform best on recognizing emotion episodes that are least ambiguous; when an expression is rated similarly both rated in the moment and at the end of day and when external raters agree with the subjective rating.

In our lab, we have made two attempts at emotion rating “in the wild.” In the first study [5], we used a cell phone with a “Mood Map” developed for a mobile heart health study [6], derived from the Circumplex model of emotion [7]. In the second study we simply used the three axis model by Merhabian [8], represented as three separate seven point scales. Figure 1 shows both ratings systems with the mean values of reported emotions superimposed. In the first study, no rating training was given to the participants and a wide range of emotion words were used. From the “free comments” section it seemed that some of these words were used colloquially or even sarcastically. The result was difficult ground truth with a high positive and high energy bias that had to be re-evaluated by third party raters to be useful [5]. In a second study, participants were given some training in emotion theory descriptors and allowed two days of journaling and feedback before the physiological sensors were introduced [9]. As a result the bias of the ratings was eliminated. Also, it was found that in this study when multiple sets of ratings were used, the machine learning result improved as the emotion instances included in the data set were successively constrained to include only those rated consistently both in the moment and at the end of the day by the participant, and then also consistently by independent raters, as illustrated in Table 1 [9]. This suggests that part of the difficulty of previous studies in emotion recognition [10] may have been that individual instances of emotion may not have been either purely felt or purely expressed as often only one rating was used.

Equally challenging to the problem of rating emotion is the problem of sensing emotion. Sensing emotion from facial expression is possible, but only if the emotion is expressed and a camera is directed at the person’s face. While this is possible in driving situations (see Figure 2) [11] and when the person is seated in front of a computer with a camera, it is difficult when the person is ambulatory [12]. Sensing emotion from speech is possible, from pitch, pitch variation, word choice, word speed, etc. [13], but only when the person is speaking. Similarly, sentiment analysis is possible from text, but only when the person authentically chooses to express emotion in text. Physiological sensing offers a constant channel to a person’s state, but physiological reactions reflect the totality of inputs to a person’s experience: physical activity, fitness level, fatigue, food intake. A sneeze might look like a startle, a cup of coffee may look like fear or anger. It is likely that not only will we have to track multiple modes of expression to capture emotion but also capture context (e.g. is the person sitting or standing? Hungry or tired?) to fully disambiguate or accurately model emotional responses. This is especially challenging in natural ambulatory environments, but new technologies are making this more possible [14, 15, 16].

Measuring driving behavior

One of the best opportunities for sensing emotion is while the person is driving. In the driving environment, the person is seated so physical activity is minimized. The car affords a constrained space into which monitoring equipment can be placed and an electrical system to power the monitoring equipment. Driving is also a place

Figure 2. Driving offers a situation where emotions can be monitored using a driver facing camera (shown in the center) and multiple external cameras (captured video shown left) and where physiological signals can be captured by a computer (right)

where many emotions are strongly felt, especially frustration, anger, and "stress" during traffic. Occasionally, emotions such as sadness or joy may occur in response to music on the radio. Driving is real life, unexpected things do occur and your life is really on the line. The driving environment offers a compromise between laboratory studies with high internal validity but limited stimuli and fully ambulatory with high external validity but difficult to interpret data.

Rating in the moment driver responses

In previous research, we have taken two approaches to rating driver responses. The first approach was fairly rigorous: driver self-assessment was used to validate the driving conditions of a planned experiment; then raters were hired to score every second of video with 10 different specific categories of stress indicators for every second of video [11]. This process was very time and labor intensive. In a second experiment, we simply sat down with participants after the driver and let them review the video, asking them to think about their relative stress on a scale of 1-7. After letting them see the entire video one time, we asked them to think about their initial stress level, then we played the video forward and asked them to indicate when they thought their level had changed. This resulted in user defined emotion intervals. In the first study, it was validated that of three driving conditions: rest (no driving), highway driving in low traffic conditions and city driving; that all physiological indicators seemed to agree that rest was least stressful, that city driving was most stressful and that highway driving was in between the two. Subject stress report across each of the conditions for each of the participants validated the study design (rest=low stress, city=high stress, highway=medium stress), then which was further supported by the sum of the second by second coding of stress metrics [11]. In the second experiment [17] we were able to conclude that GSR correlated with driver stress on an episodic basis. For this analysis, due to GSR baseline issues we have encountered in the past, [5] we used mean relative maxima for the results in the left hand side of the Table 2 and relative differences in mean values between episodes corresponding to differences in ratings for the results in the right side of Table 2. This highlights, for example that strong *changes* in state, such as the one that occurred between the "3" rated episode starting at 410 seconds and the "7" rated episode starting at 510 seconds also corresponded to the highest change in the GSR. These types of ratings give us a method of rating driver responses "in the moment" without having to distract the drivers while they are driving and give us a short term view of driver behavior. Ideally when sensing methods become trivially burdensome to participants and ratings can be made automatically from video indicators, these detailed snapshots of behavior can be aggregated across users and across time.

Rating longer term driver behavior

In another study, we took a different look at driver behavior, this time using only GPS data from their car and an application monitoring program on their phones [18]. The purpose of this study was to see how and when people used their phones while driving, but the study revealed far more than that. We were able to infer people's routine drives and identify exceptional events, sometimes even when these events did not seem exceptional to the people themselves. For example, Figure 4 shows a Google Earth plot of the GPS and cell

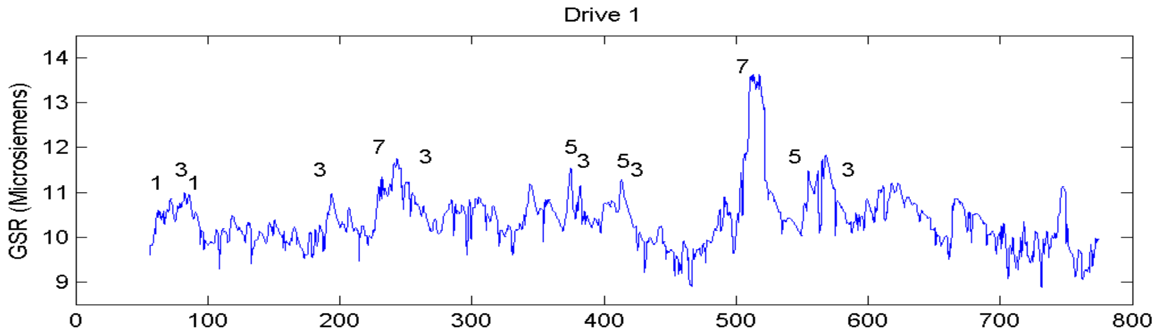


Figure 3. A plot of galvanic skin response (GSR) during the drive with the participant's ratings from the video superimposed.

Table 2. A comparison of participant ratings and relative GSR maximums, followed by a comparison of the mean changes in GSR corresponding to the change in ratings between successive user-defined segments.

Rating	1	3	5	7	Δ Rating	-4	-2	2	4
GSR Max	0.3	0.6	1.1	3.2	Δ Mean (μS)	-0.58	-0.24	0.26	2.0

phone activity for one of our participants. We had identified this as an anomalous event due to the amount of cell activity while driving and the erratic path. When the participant was first asked about this data, she recalled that

Figure 4. A Google Earth plot of driving activity overlaid with phone activity. The colors of the trace indicate speed: red (under 5km/h) , orange (5-10km per hour), blue (20-50km per hour), dark green (50-70) km per hour and light green (>70kph). The white text indicates the app that came into focus at the time indicated by the pinpoint near the annotation.

at this point her friend had contacted her to report that there was a gang of "pirates" in the area. This is a large group of thieves who move from car to car and rob them. She was driving to avoid them and at one point heard a gunshot. She finally felt safe at a restaurant in the lower left of the graph and called her friend to let her know. By automating anomalous event detection, we might intelligently identify such interesting events for labelling and analysis.

Conclusions

Although the methods of measuring behavior presented here are different, they all rely on the interplay between sensor data and human interpretation of experience. For the studies using physiological monitoring we were limited in the number of subjects we could have and the length of the study because of the expense and fragility of the research grade sensing systems and the labor intensive process of collecting and interpreting annotations. For the long term driving behavior study, we were able to run multiple participants in different parts of the world for a month at a time the monitoring equipment was low cost and easily, invisibly integrated into their daily lives. The GPS sensor took up no space in their car, and the app took up very little space on their phones, and since we recruited only participants who already owned Android phones, we eliminated this expense. We are on the cusp of seeing new possibilities for measuring behavior as sensors begin to be integrated into our clothes and daily lives. Products like FitBit, Nike Fuel, Basis, Om shirt, BodyMedia and Polar HRMs are making monitoring easier, however unlike GPS sensors it is still difficult to access and share data from these devices.

We look forward to the next generation of commercial sensing that leverages the platforms that are already part of our daily lives: our computers, homes, cars, phones, clothing and shoes that also allow us full access to our own data; to see for ourselves and to make our own inferences or to allow us to share our data with others in a way that allows us to understand and control how much we share. When we can aggregate data across multiple sources we will be able to see behavior from different perspectives, in contexts and across millions of people. With this kind of vast data set, different kinds of machine learning algorithms can be tried to make sense out of data and ratings we can then start to identify trends and groupings that have never before been imagined or that we have been able to isolate and test because of the current constraints on equipment and budgets. We must encourage the next generation of sensors to allow access to data so that we can all participate in finding these insights.

References

[1] Ekman, P. (1999). Basic Emotions. *In Handbook of Cognition and Emotion.* Chichester, West Sussex, England: John Wiley and Sons, Ltd.

[2] Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion,* ***9***(1), 83-91.

[3] Oatley, K., Duncan, E. (1994). The Experience of Emotion in Everyday Life. *Cognition and Emotion,* ***8***(4), 369-381.

[4] Fahrenberg, J., Myrtek, M. (2001). *Progress in ambulatory assessment*. Seattle, WA, USA: Hogrefe & Huber Publishers.

[5] Healey, J., Nachman, L., Subramanian, S., Shahabdeen, J., Morris, M. (2010). Out of the lab and into the fray: Towards modeling emotion in everyday life. *Lecture Notes in Computer Science (Pervasive Computing),* **6030**, 156-173.

[6] Morris, M., (2007). Technologies for heart and mind: New directions in embedded assessment. *Intel Technology Journal,* ***11(1)****,* 71-75.

[7] Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology,* ***39(6)****,* 1161-1178.

[8] Russell, J.A., Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality,* ***11(3)****,* 273–294.

[9] Healey, J. (2011). Recording affect in the field: Towards methods and metrics for improving ground truth labels. *Lecture Notes in Computer Science (Affective Computing and Intelligent Interaction),* ***6974***, 107-116.

[10] van den Broek, E.L., Janssen, J.H., Westerink, J.H.D.M., Healey, J.A. (2009). Prerequisites for Affective Signal Processing (ASP). In *Biosignals 2009: Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing,* 14-17 January, Porto, Portugal. pp. 426-433. Setubal, Portugal: INSTICC.

[11] Healey, J., Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological

sensors. *Transactions on Intelligent Transportation Systems*, ***6***(2), 156-166.

[12] Teeters, A., el Kaliouby, R., Picard, R.W. (2006). Self-Cam: Feedback from what would be your social partner. In *Proceedings of SIGGRAPH: The 33rd International Conference on Computer Graphics and Interactive Techniques*. 1-3 August, Boston, MA, USA. Article no. 138. New York, NY, USA: ACM.

[13] Schuller, B., Steidl, S., Batliner, A. (2009). Sensing emotion: The INTERSPEECH 2009 emotion challenge. In *Proceedings of the 10th International Conference on Spoken Language Processing (INTERSPEECH)*. 6-10 September, Brighton, UK. pp. 312-315. ISCA.

[14] Westerink, J., Ouwerkerk, M., de Vries, G.-J., de Waele, S., van den Eerenbeemd, J., van Boven, M. (2009). Emotion measurement platform for daily life situations. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, 10-12 September, Amsterdam, The Netherlands. pp. 217-222. IEEE Press.

[15] Kocielnik, R. (2014). LifelogExplorer: A tool for visual exploration of ambulatory skin conductance measurements in context. In *Proceedings of Measuring Behavior 2014, 9th International Conference on Methods and Techniques in Behavioral Research.* 27-29 August, Wageningen, The Netherlands. URL: http://www.measuringbehavior.org/files/2014/Proceedings/Kocielnik,%20R.D.%20-%20MB2014.pdf [Last accessed on 7 November 2014].

[16] Antony, J., Sharma, K., Castellini, van den Broek, E.L., Borst, C. (2014). Continuous affect state annotation using a joystick-based user interface. In *Proceedings of Measuring Behavior 2014, 9th International Conference on Methods and Techniques in Behavioral Research.* 27-29 August, Wageningen, The Netherlands. URL: http://www.measuringbehavior.org/files/2014/Proceedings/Antony,%20J.%20-%20MB2014.pdf [Last accessed on 7 November 2014].

[17] Healey, J., Theocharous, G., Kveton, B. (2010). Does my driving scare you? In *Adjunct Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010)*, 11-12 November, Pittsburgh, PA, USA. p. 14. URL: http://www.auto-ui.org/10/adjunctproceedings/p14.pdf [Last accessed on 7 November 2014].

[18] Zafiroglu, A., Healey, J., Plowman, T. (2012). Navigation to multiple local transportation futures: Cross-interrogating remembered and recorded. In *AutomotiveUI '12: Proceedings of Automotive User Interfaces and Interactive Vehicular Applications*, 17-19 October, Portsmouth, NH, USA. pp. 139-146. New York, NY, USA: ACM.

Eyeblink conditioning:
A paradigm to study associative learning across species

John F. Disterhoft, Shoai Hattori, Taejib Yoon, Carmen Lin and Craig Weiss

Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
jdisterhoft@northwestern.edu

Introduction

The trace eyeblink conditioning paradigm, a forebrain dependent associative learning task, was introduced. This learning paradigm becomes forebrain dependent when a stimulus-free trace interval intervenes between the conditioned stimulus and the unconditioned stimulus. The requirement that the conditioned stimulus be remembered over time makes the task forebrain-dependent. This paradigm was initially developed in humans and subsequently used to study the substrates of learning in a variety of mammalian species including rabbits, rats and mice. Our presentation focused on the use of eyeblink conditioning, when combined with in vivo single neuron recording, to study the involvement of the CA1 hippocampus and prefrontal cortex in the acquisition and consolidation of this learned task. All experimental procedures carried out using animal subjects are approved by the Northwestern University Institutional Animal Care and Use Committee. The minimum number of animal subjects required to obtain statistically meaningful results are used.

The eyeblink conditioned response: A series of three studies

The first series of studies reviewed were those addressing the role of the CA1 hippocampus in eyeblink conditioned response elicitation during overtraining and, most importantly, after one month of memory consolidation had occurred. A contentious point in memory research is whether or not the hippocampus plays a time-limited role in the consolidation of declarative memories. A widely held view is that declarative memories are initially encoded in the hippocampus, then transferred to the neocortex for long-term storage. Alternate views argue instead that the hippocampus continues to play a role in remote memory recall. These competing theories are largely based on human amnesic and animal lesion/inactivation studies. Important animal lesion studies supporting the standard consolidation theory have been done in rabbits using trace eyeblink conditioning. However, *in vivo* electrophysiological evidence supporting these views is scarce. Given that other studies examining the role of the hippocampus in remote memory retrieval using lesion and imaging techniques in human and animal models have provided mixed results, it was particularly important to gain insight at the *in vivo* electrophysiological level. We summarized hippocampal single-neuron and theta activity recorded longitudinally during acquisition and remote retrieval of trace eyeblink conditioning. Results from conditioned rabbits were compared to those obtained from yoked pseudo-conditioned control rabbits. Results revealed continued learning-specific hippocampal activity one month after initial acquisition of the task. Our findings yield insight into the normal physiological responses of the hippocampus during memory processes and provide compelling *in vivo* electrophysiological evidence that the hippocampus is involved in both acquisition and retrieval of consolidated memories.

The second series of studies reviewed addressed the role of subregions of the prefrontal cortex in the acquisition and consolidation of the trace eyeblink conditioned response. The medial prefrontal cortex (mPFC) has been studied for its role in various cognitive functions, but the roles of its subregions remain unclear. We performed tetrode recordings simultaneously from prelimbic (PL) and rostral/caudal anterior cingulate (rACC/cACC) subregions of the rabbit medial prefrontal cortex to understand their interactions during learning and remote memory retention for whisker-signaled trace eyeblink conditioning. cACC neurons exhibited an innate response to the conditioning stimulus that rapidly decreased across sessions, suggesting an attentional role for facilitating conditioned stimulus-unconditioned stimulus associations. rACC neurons from conditioned rabbits exhibited robust responses to the conditioned stimulus that decreased within each session, possibly evaluating its emotional salience. PL neurons exhibited robust persistent activity during the trace interval during tests of remote memory retention, suggesting its involvement in retrieval and execution of a consolidated response. Mechanistically, conditioning was associated with a greater percentage of persistently responsive neurons than neurons from pseudoconditioned control rabbits, and responses differed significantly between trials with and without conditioned responses. Collectively, these responses reflect a functional reorganization of neural activity within the prefrontal network from an attentional mode in the cACC to one in PL that orchestrates the retrieval and execution of the learned response after memory consolidation has occurred.

The final series of studies reviewed demonstrated a new behavioral protocol for training head restrained mice in trace eyeblink conditioning. Whisker vibration is used as a conditioned stimulus and periorbital shock is used as an unconditioned stimulus in these experiments. Mice show no signs of stress in this situation, in which they are allowed to run on a freely rotating cylinder during training. This freedom to run seems to relieve stress in mice, even when their head is fixed. This procedural modification is important since mice do show stress when put in a situation in which their body is totally restrained. The conditioned mice show acquisition of the eyeblink conditioned response during 10 training sessions; there is no change in eyeblink responses to the whisker vibration conditioned stimulus in pseudo-conditioned control mice presented with unpaired presentations of the whisker vibration and periorbital shock. This behavioral paradigm is being developed for use in experiments that will do single neuron recording and/or in vivo 2-photon calcium imaging of single neurons in vivo in conscious mice. Using mice will also facilitate experiments using molecular genetic and optogenetic approaches to analyze systems and cellular mechanisms of learning in mice.

Conclusions

Our presentation focused on the utility of the eyeblink conditioning associative learning paradigm as a well-controlled behavioral model system with which to analyze the cellular and systems mechanisms of associative learning and memory storage in mammalian brain. The parameters of learning tend to be relatively homogeneous across animals, facilitating these studies. This model system also allows including a powerful pseudo-conditioning control group that receive the same number of conditioned and unconditioned stimuli in an unpaired fashion. The pseudo-conditioning group controls for the possibility of behavioral sensitization to the presentation of the conditioned or especially the unconditioned stimuli. Eyeblink conditioning can be combined with permanent lesions, inactivation, in vivo single neuron recording, and ex vivo biophysical or biochemical measurements to analyze the systems, cellular and/or molecular mechanism of associative learning in mammalian brain.

Acknowledgments

This work was supported by National Institute of Mental Health grants F31MH099769 (SH), MH47340 (JFD), National Institute of Aging grants 5T32AG020418-10 (SH), 5R37AG08796 (JFD), 5R01AG017139 (JFD) and National Institute of Neurological Diseases and Stroke grant NS059879 (CW).

Symposiums

The effects of social communication:
a research study on neuroscientific techniques application

M. Mauri[2], A. Ciceri[1,2], G. Songa[1,2], F. Sirca[2], F. Onorati[2,3] and V. Russo[1,2]

[1] Institute of Consumption, Behavior & Communication, IULM University of Language and Communication of Milan, Italy vincenzo.russo@iulm.it

[2]Behaviour and Brain Lab, IULM University of Language and Communication of Milan, Italy brainlbab@iulm.it

[3] Dipartimento di Elettronica, Informatica e Bioingegneria, Politecnico of Milan, Milan, Italy francesco.onorati@polimi.it

Introduction

The effectiveness of social advertising is still an open issue in the scientific community. On one side the evaluation of advertising efficacy is still a big challenge to accomplish in terms of scientific measures. On the other side, the social advertising is often based on emotional appeals, in order to capture and affect audience attention, and the attempt to measure emotional reactions in a scientific way is still debated by scientists as well. Nowadays the evaluation of advertising is laying on traditional techniques, the so called "explicit self-report measures", based on direct questions to subjects exposed to advertisement. The limit of these traditional techniques is the potential overlap (and confusion) of what is felt by subjects by means of their rational appraisal with what is really perceived by them from an emotional standpoint [1,2].

Today the technological improvements and neuroscientific findings provide an important knowledge that might contribute to develop new kind of measures based on neuro and psycho-physiological activity correlating with affective reactions during the advertising exposure. In comparison to "direct self-report measures", these techniques are based on *indirect* measures ABLE to provide more complete information about subjects reactions and reduce the potential contributes from affective bias.

Purpose and methodology

The present work presents the results from a scientific study about a social spot (lasting 50 seconds) aimed to sensitize public opinion about discrimination made by men against women.
The goal of the research was to evaluate the different effects played by the spot on male subjects in comparison to female ones. Previous research already showed important differences between males and females regarding social communication, in particular about the level of engagement [3] and the cortical activation measured by electroencephalography [4,5].

Aside traditional methods based on self questionnaires, some neuroscientific techniques were also applied, as eye-tracking recordings with a RED 250 system from SMI [6] synchronized with electroencephalography (EEG) monitoring from Emotive Epoc headset [7,8] while watching the spot. In addition to these techniques has been used the automatic quantitative analyses of Emotional facial expressions computed by FaceReader 5, a software from Noldus [9,10].
60 subjects, 30 males and 30 females, have been enrolled in the study, they were all students from IULM University. After filling up consent forms about the research purposes and ethical issues, they have been asked to sit down in front of the computer in order to watch the spot while all their eye-gaze, facial expressions and EEG brain waves were monitored.

Results

All subjects have filled up a self-report questionnaire after watching the spot . According to results, the females judged the spot better, with an average of 4,84 over a Likert scale [11] ranging from “1 = I did not like the spot at all” to “7 = I liked the spot very much” in comparison to males (with the average of 5,56). However, the difference has been not statistically significant (t-test: $p \geq 0,05$). Moreover, the results show a qualitative difference enabling to understand that females liked the spot more than males, but without any additional information about the affective and cognitive processes that leaded to this preference. For this reason, the qualitative result has been integrated with further information from neuroscientific tecniques application.
The results from the automatic quantitative analyses of facial emotional expressions of all subjects while watching the spot show that on average the whole sample reacted with a neutral face (dominating on all other facial emotional expressions – fear, anger, sadness, happyness, surprise and disgust) for the 63% of the whole exposure, on a range between 0 and 100%. The second predominant facial reaction has been “Sad” for the 21 % on average of the whole esposure of all subjects. We can conclude that the main facial emotional reaction of all subjects (aside “neutral” condition) has been “sad” on average. (See the box plot graph showed in Figure 1).

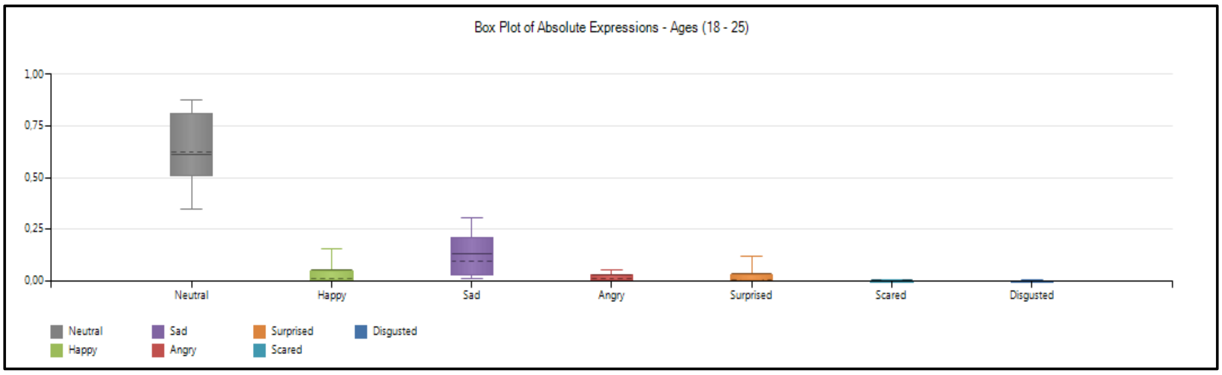


Figure 1. Face Reader outputs for the whole sample of 60 subjects. On the “x” axe, from left to right, emotional conditions automatically computed by the software: Neutral, Happy, Sad, Angry, Surprised, Scared and Disgusted. On the “y” axe, the values are expressed between “0” and “1” (0= probability that the face is showing one of the 7 conditions is equal to 0%; while 1= probability that the facial expression is showing one of the 7 conditions is equal to 100%).

If the results are analyzed according to male and females subgroups, it is possible to observe that results are not the same. Females felt emotions of “happiness” (0.05=5%) and “surprise” (0.04=4%) (even if in small percentage, the difference with males is significant, t-test: $p \leq 0.05$) (see the graphs in Figure 2; in purple the happiness, in orange the surprise).
Instead, in the subgroup of male subjects, it is possible to see that they reacted, aside “neutral” and “sad”, with some facial expressions detected as “angry” (0.03 = 3%), significantly different from females (who reacted with an “angry” face with 0.01=1% - t-test: $p \leq 0.05$), as shown in the graph below (see the graph in Figure 2; neutral in grey, angry in red).

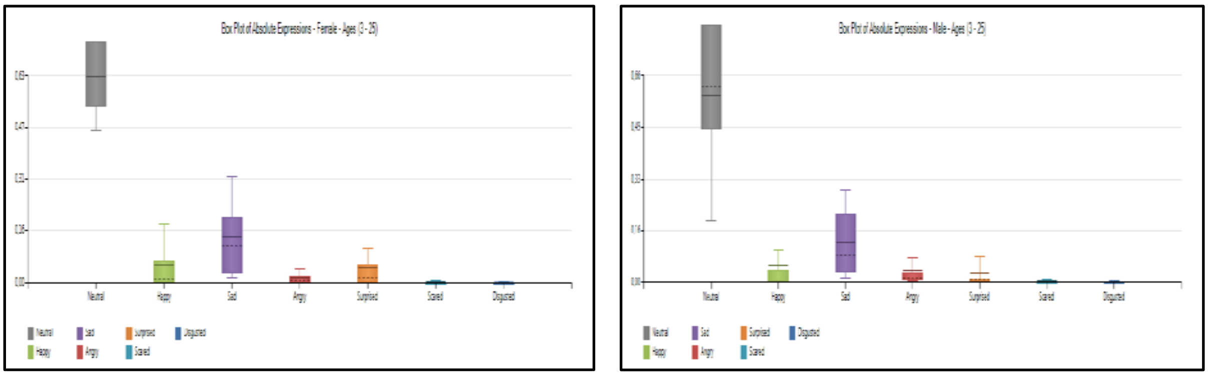


Figure 2. On the left, Face Reader outputs for the female sample of 30 subjects. On the right the output for the male sample.

Considering the automatic analyses from the Face Reader software about the level of valence (the probability, btween "+1.00" – "positive" 100% and "–1.00" – "negative" 100 % - whereas "0.00" is equal to "neutral" condition) of the face patterns there are significant differences (t- test: $p \leq 0.05$) between males and females: males showed on average a **negative valence** of the face that is almost the double (-0.18) in comparison to famels (-0.10).

Aside the data concerning the facial expression analyses, the data from electroencephalography (EEG) were also monitored while the 60 subjects watched the spot. Data were computed by the Affective Suite software from Emotiv that automatically processes all the data, that is providing 3 indicators: a) "engagement", the general cortical activation elicited by a stimulus or a situation; in a very broad way, this pattern of cortical activation is characterized by a predominance of beta waves activity, more associated with increased cognitive and/or affective mental states [12,13] with a decreased activity of Alpha waves, more active in relaxation situations [14,15]; b) "Excitement short term", the cortical activation elicited in a time window of few seconds after a stimulus, characterized by the predominant activity of sympatetic branch of the autonomous noervous system in comparison to the parasympathetic activity branch), c) "frustration", the cortical activity associated with cognitive and affective process while trying to cope with a situation characterized by negative emotional states. Further analyses about classic bandwiths (as Alpha waves, Beta waves, etc.) will be computed and analysed in future paper works. Data were then cleaned from motion artifacts and outliers, and averaged across all male and female subjects, along the whole exposition to the spot.

Examining the results about "Engagement", it is possible to see (Figure 3) how the females subjects (in pink) are almost always showing higher values in comparison to male ones (in blue). Females, on average, showed an engagement value higher (0,65=65%) in comparison to males (0.62=62%); however, this can be considered such a trend only, as difference were not significant (t-test: $p \geq 0.05$).

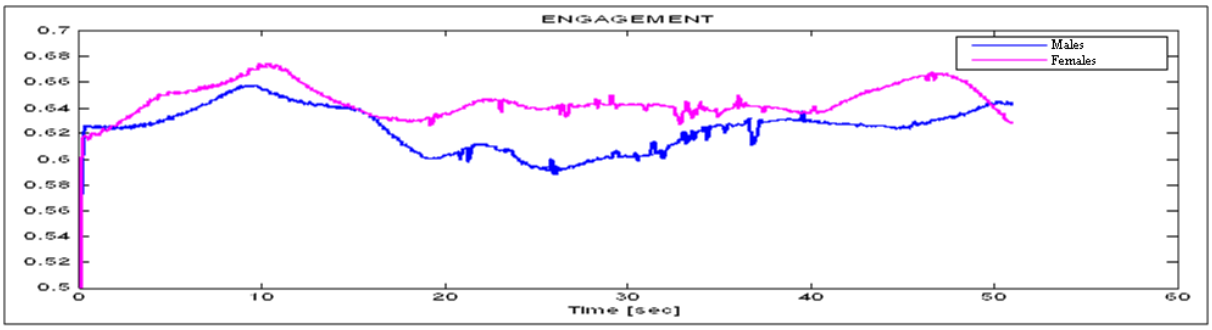


Figure 3. "Engagement" index for males and females while watching the social spot; on "x" axe time is expressed in seconds, while on the "y" axe the level is expressed in percentage (from 0.00=0% to 1.00=100% of the level of "engagement").

Considering the “frustration” index, males showed (Figure 4) an higher “frustration” index (0.48=48%) in comparison to females (0.44=44%). However also this result has to be taken into account as a trend, since there are not significant differences (t-student test: p≥0.05)

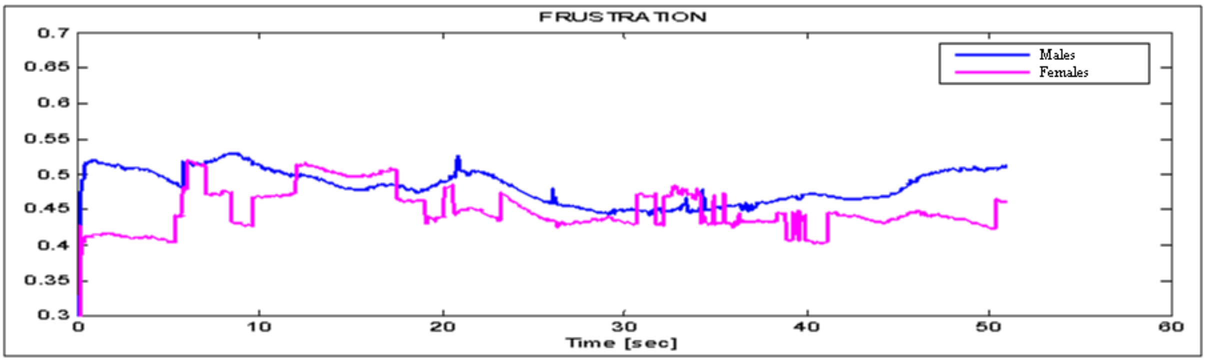


Figure 4. “Frustration” index level for males and females while watching the social spot; on “x” axe time is expressed in seconds, while the “y” axe the level is expressed in percentage (from 0.00=0% to 1.00=100% of the level of “frustration”).

Last, looking at the index “Excitement short term” has been considered and analyzed it is possible to notice that males are always showing higher values in comparison to females ones (Figure 5). The difference is significant (t- test: p≤0.05).

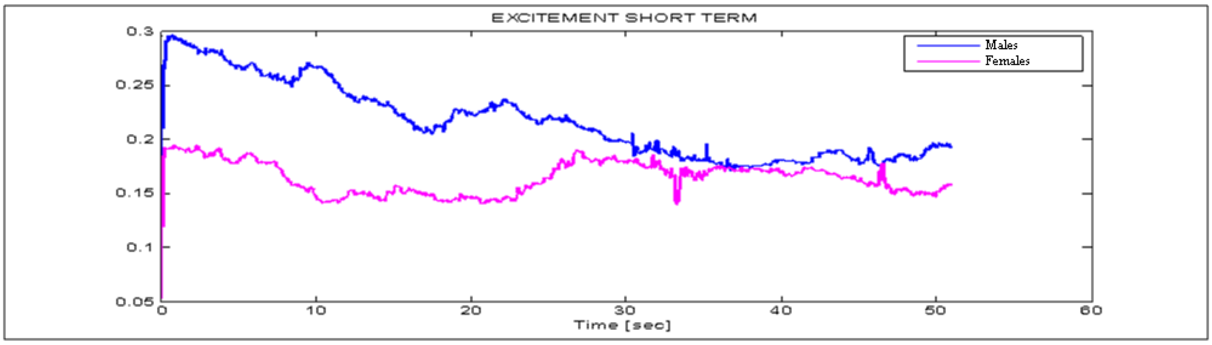


Figure 5. “Excitement short term” index from EEG data for males and females while watching the social spot; on “x” axe time is expressed in seconds, while the “y” axe the level is expressed in percentage (from 0.00=0% to 1.00=100% of the level of “Excitement short term”).

Considering now the data from the Eye-tacking system, the synchronization between the EEG indicators and the pointing of the gaze is showed by Figure 6. In the upper part of Figure 6 some frames from the spot are showed, with correspondant values from EEG Affective Suite indicators from EEG (in the lower part of the figure). The peaks of red line (corresponding to the average index of “engagement” across all 60 subjects) are linked with scene from the spot where male charachters are showing negative social behaviors (such as writing on the white space of the advertising bills of the social campain in favour of wemen) against women. Moreover, the eye-fixations analyses showed different visual behavior between males and females. On one side, males looked on average more at male characters faces (in the social spot considered, the only male characters enrolled were showing a stereotyped social behavior against women) in comparison to female participants. Males, over 50 seconds of the whole spot exposition, spent on average almost 5 seconds (4916 milliseconds) in looking at male characters faces (appearing for a total of 7 seconds over the 50 seconds), while females spent on average almost 4 seconds (4025 milliseconds in total). The difference between males and females gaze behavior is significant (t-test: p≤0.05). On the other side, female subjects looked on average more at female charachter faces (28 seconds in total) in comparison to male participants (26 seconds in total), however the difference is not significant (t- test: p≥0.05)

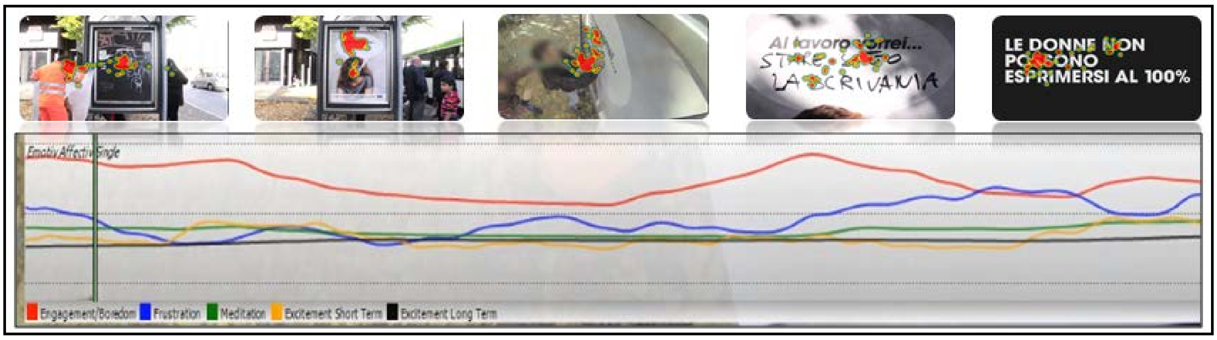


Figure 6. In the lower side of the picture, EEG indicators are showed: red line is “Engagement” trend, blue line is “Frustration” index, while “Excitement shor term” is the yellow line. In upper part of the figure, the correspondant frames of the spots.

Conclusions

In conclusion, according to self-report results, it is possible to claim that female subjects liked the social spot more than male ones. Self-reports represent one of the traditional ways to evaluate the effectiveness of communication in consumer psychology. However, the application of some neuroscientific techniques allows to go beyond the traditional methods in order to understand in a more complete way the consumer reactions. For instance, from facial expression analyses, it is possible to claim that females showed less negative emotions in comparison to males, who watched more the male characters enrolled in the spot (see the data from the eye-tracking) who were doing bad social actions against women. Combining results from facial expression and from eye-tracking, it is possible to claim that male participants identified themselves more with males characters who were showing negative social behaviors against women. This is the reason why they reported more negative emotional expressions by means of their face. In addition, also the EEG results showed a stronger activation about “frustration” and “excitement short term”, supporting the claim that male subjects reacted with a stronger negative emotional reaction in comparison to female subjects, who were possibly feeling better, thus showing less negative emotional reactions as they were watching a social communication rejecting stereotyped negative social behavior against women usually made by men as depicted by the spot.

Aside the significant results that allows to understand why females appreciated the spot better in comparison to males, the aim of this work is to show also how the integration of traditional and innovative techniques from neuroscience can provide a wider range of information enabling a wider understanding of subjects’ evaluation processes, as the empirical analyses of quantitative data showed in the results section of this research work. Even if the positive results must to be taken into account with caution, due to the pioneering application of neuroscientific procedures to the field of communication, implications, limits and advantages are discussed from a methodological point of view for further future research questions.

References

[1] Schachter, S., Singer, J. (1962). Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review*, ***69***, 379-399.

[2] Kahneman, D., Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, **47**(2), 263-291.

[3] Berger, I., Cunningham. P., Kozinets R. (1999). Consumer Persuasion Through Cause-Related Advertising. *Advances in Consumer Research*, ***26***, 491-497.

[4] Pradeep, A. (2010). *The Buying Brain: Secrets for Selling to the Subconscious Mind*, John Wiley & Sons.

[5] Vecchiato, G., Astolfi, L., De Vico F., Toppi J., Aloise F., Bez F., Daming W., Wanzeng K., Jounging D., Cincotti F., Babiloni F. (2011). On the Use of EEG or MEG Brain Imaging Tools in Neuromarketing Research, *Computational Intelligence & Neuroscience*. 1-12.

[6] Pasqualotti, L., & Baccino, T. (2014). Online advertisements: how are visual strategies affected by the distance and the animation of banners? *Frontiers in Cognition*, **7**.

[7] Bagdžiūnaitė, D., Nassri, K., Clement, J., Ramsøy, T.Z. (2014). An added value of neuroscientific tools to understand consumers' in-store behaviour. *EMAC 2014*.

[8] Khushaba R. N., Wise C., Kodagoda S., Louviere J., Kahn B. E., Townsend C. (2013). Consumer Neuroscience: Assessing the Brain Response to Marketing Stimuli Using Electroencephalogram (EEG) and Eye Tracking, *Expert Systems with Applications*.

[9] Lewinski, P., Fransen, M., Tan, E. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. *Journal of Neuroscience, Psychology, and Economics*, **7**(1).

[10] Ekman, P. (1992). Facial expressions of emotion: New Findings, New Questions. *Psychological Science*, **3**, 34-38.

[11] Likert, R. (1932). Technique for the measure of attitudes. *Archives of psychology*, **22**.

[12] Scotti S., Mauri M., Barbieri R., Brown E.N., Cerutti S., Mainardi L., Villamira L., (2006) "Automatic quantitative evaluation of Emotions in e-Learning Applications", in *the IEEE International Conference of the Engineering in Medicine and Biology Society*, (New York, August 30 – 3 September 2006).

[13] Mauri M., Magagnin V., Cipresso P., Mainardi L., Brown E. N., Cerutti S., Villamira M., Barbieri R., (2010). "Psychophysiological signals associated with affective states", *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (Buenos Aires, August 31 – 4 September, 2010).

[14] Myers J. E., Young J. S., (2012). Brain Wave Biofeedback of Integrating Neurofeedback in Counseling, Journal of Counseling & Development, **9**, 20-28.

[15] Haarmann H.J., George T., Smaliy A., Dien J., (2012). Remote Associates Test and Alpha Brain Waves. *Journal of Problem Solving*, **4**, 66-93.

Integration of traditional and innovative methods in studying advertisements effectiveness via paper, tablet and website: a neuromarketing experiment

V. Russo[1,2], M. Ferraresi[1], A. Ciceri[1,2], G. Songa[1,2], F. Onorati[2,3], F. Sirca[2] and M. Mauri[2]

[1] Istituto di Consumi, Comportamenti e Comunicazione, Libera Università di Lingue e Comunicazione IULM of Milan, Milan, Italy
vincenzo.russo@iulm.it, mauro.ferraresi@iulm.it
[2] Behaviour and Brain Lab, Libera Università di Lingue e Comunicazione IULM di Milano, Milan, Italy
brainlab@iulm.it
[3] Dipartimento di Elettronica, Informatica e Bioingegneria, Politecnico of Milan, Milan, Italy
francesco.onorati@polimi.it

Introduction

In the last decades the availability of portable devices as well as of internet accesses provided such great changes in people's habits in taking advantage of media contents. Nowadays people are exposed to advertising through several media, such as newspapers, websites, and mobile/tablet applications . To establish which media might be more effective in terms of ads memorization is an ardous challange, but the answer has a great importance for strategic media planning of advertising communication. Some studies have already shown that media via paper are more effective than media via internet and websites in promoting the recall and the recognition of the ads [1,2]. However, the scientific debate about the effectiveness of some media over the others is still an open issue.

Aim

The aim of this work is to evaluate which communication media between newspapers, websites and mobile applications is able to induce the higher ads memory performances. For this purpose, we have applied both traditional methods, based on self-reports/interviews, and neuromarketing techniques [3], such as Eye-Tracking (ET) recordings synchronized with Electroencephalographic (EEG) signals.

Design/methodology

The present study was carried out at the Behavior and Brain Lab (IULM University, Milan) with the participation of 72 subjects equally distributed in terms of age, gender, socio-economic level and habits in the use of communication media. Three subgroups of 24 subjects, one for each experimental condition (printed newspaper, website and mobile application on iPad), were arranged.

The experiment was divided into two phases: firstly, the subjects were asked to enojoy the reading as they were at home an Italian newspaper through three different media (i.e. print version, mobile version on iPad and website version), each one including ads of the same brands.
In the second phase every subject has been interviewed for an implicit and explicit [4] memory task recognition over 50 ad flyers [5,6], half of them included in the respective medium during the first phase of the experiment, while the remaining ones served as distractors.

It is important to point up that for each medium the ads were always the same. On the other hand, the newspaper articles were the same in the print version and for the mobile version, while for the website version they could change. This choice was made to respect the ecologial conditions of exploring website versions of newspapers. At the end of the memory recognition task, subjects were asked to fill in a self-report assessment about their brand knowledge [7] to verify wheather the memory task was affected or not by previous experiences (such as holding or willing to purchase one of the products belonging to the brands presented by the target ads used during the experiment).

The Emotiv EPOC electroencephalography (EEG) headset was synchronized with SMI Eye-Tracking glasses for the print version and the mobile version groups, while for the website version group the Emotiv Epoc EEG headset was syncronized with the SMI RED250 Eye-Tracking (ET) [8,9,10]. The analisys of eye movements is a good source of relevant informations. The fixation of an object, in fact, indicates that the subject paid attention to that, and the length is a good indicator of the amount of interest [11,12]. Among the different areas of research that have benefited from the measurement of gaze behavior, there is the study of advertising [13,14,15,16], as the level of attention paid to an advertising influence the perception of the message [17] that can influence the attitude and consequently the behavior [18]. ET and EEG recordings were monitored also during the memory recognition task.

Results

Results show that print (paper) and mobile (iPad) versions obtained significant higher values in memory performances (t-student test: $p \leq 0.05$). Mobile (iPad) version had a higher memory rate than print (paper) version even if not in a statistically significant way.
Moreover, as showed by Figure 1, on average the total eye fixation time (computed by eye-tracking SMI BeGaze software) on ads was lower in the website version than both tablet version (iPad) and printed version (paper), according also to the results showed by some previous studies wich found the “banner blindness” phenomenon [19].

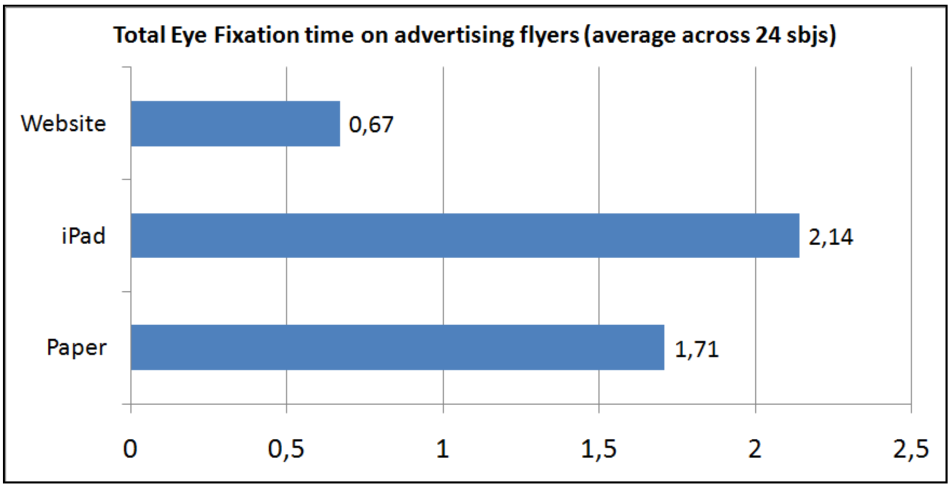


Figure 1. The graph shows the average values (expressed in seconds on the “x” axe, averaged across all 24 subjects for each media) about the total time of eye fixations on advertising flyers for the 3 different media exposition: paper, iPad and website (on the “y” axe).

EEG analyses shows that the Frustration Index from Emotiv Affective Suite is significantly lower in the website version with respect to both paper version and tablet version (t-student test: $p \leq 0.05$), while there were no significant differences (t-student test: $p \geq 0.05$) between paper version and tablet one (see Figure 2).
Moreover, the Frustration Index correlates significantly with memory performances (correlation test: $p \leq 0.01$) and this finding is consistent with some empirical evidence about the relation between emotion and memorization [20,21,22,23].

Conclusion

Aside the findings presented in the results section, which should be taken into account with caution due to the small sample size enrolled in this study, this work shows that the integration of traditional methods (i.e., self-reports, interviews and traditional memory task recognition) and innovative techniques (such as eye-tracking recordings synchronized with EEG monitoring of brain activities) might provide a more complete information and understanding of user memorization phenomena about ads memorization. Limits and advantages of the presented neuromarketing methodologies will be deepen in further works.

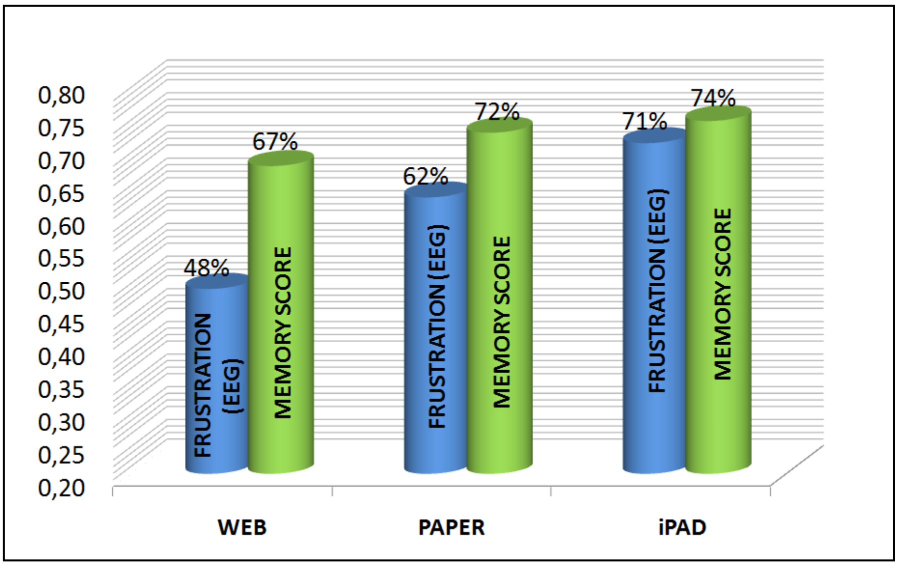


Figure 2. The graph shows the average values (percentages on the "y" axe, across all 24 subjects for each media) about the total memory task scores (in green) and EEG frustration levels (in blue) about advertising flyers for the 3 media: paper, iPad and website (on the "x" axe).

References

[1] Sundar, S. S., Narayan, S., Obregon, R., & Uppal, C. (1998). Does Web advertising work? Memory for print vs. on line media. *Journalism & Mass Communication Quarterly*, **75** (4), 822–835.

[2] Bezjian-Avery, A., Calder, B., & Iacobucci, D. (1998). New media interactive advertising vs. traditional advertising. *Journal* of *Advertising Research* , **38** (4), 23–32.

[3] Plassmann, H., Ramsøy, T. Z., & Milosavljevic, M. (2012). Branding the brain: A critical review and outlook. *Journal of Consumer Psycholog*y, **22** (1), 18–36.

[4] Shapiro, S., Krishnan, H.S. (2001). Memory-Based Measures for Assessing Advertising Effects: A Comparison of Explicit and Implicit Memory Effects. *Journal of Advertising*, **30**(3), 1–14.

[5] Keller, K. L. (1987). Memory factors in advertising: The effect of advertising cues on brand evaluations. *Journal of Consumer Research*, **14**, 316–333.

[6] Braun, K. A. (1999). Postexperience advertising effects on consumer memory. *Journal of Consumer Research*, **25** (4), 319–334.

[7] Kent, R. J., Allen, C. T. (1994). Competitive interference effects in consumer memory for advertising: the role of brand familiarity. *Journal of Marketing*, **58** (3), 97–105.

[8] Esfahani, E. T. and Sundararajan, V. (2011). Using Brain-Computer Interfaces to Detect Human Satisfaction in Human-Robot Interaction. *International Journal of Humanoid Robotics*, **8** (1), 87–101

[9] Khushaba, R. N., Wise, C., Kodagoda, S., Louviere, J., Kahn, B. E., Townsend, C. (2013). Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking. *Expert Systems with Applications*, **40** (9), 3803–3812.

[10] Krober-Riel, W. (1984). Effects of emotional pictorial elements in ads analized by means of eye movement monitoring. *Advances in Consumer Research*, **11**. 591–596.

[11] Hollingworth, A., & Henderson, J. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology,* **127**, 398–415.

[12] Pieters, E., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *Journal of Marketing,* **68**, 36–50.

[13] Mackenzie, S. B. (1986). The role of attention in mediating the effect of advertising on attribute importance. *Journal of Consumer Research*, **13**, 174–195.

[14] Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Safety Science, **41**, 427–443.

[15] Pieters, R., & Wedel, M. (2007a). Informativeness of eye movements for visual marketing. In P.Wedel,&R. Pieters (Eds.), Visual marketing: From attention to action (pp. 43–71). Hillsdale, NJ: Erlbaum.

[16] Pieters, R., & Wedel, M. (2007b). Goal control of attention to advertising: The yarbus implication. *Journal of Consumer Research*, **34**, 224–233.

[17] Rayner, K., & Castelhano, M. S. (2007). Eye movements during reading, scene perception, visual search, and while looking at print advertisements. In P.Wedel, & R. Pieters (Eds.), Visual marketing: From attention to action (pp. 9–42). Hillsdale, NJ: Erlbaum.

[18] Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science*, **19** (4), 297–312.

[19] Benway, J. P. & Lane, D.M. (1998). Banner blindness: Web searchers often miss 'obvious' links. *Internetworking*, **5**.

[20] Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 379–390.

[21] Christianson, S. A., & Fallman, L. (1990). The role of age on reactivity and memory for emotional pictures. *Scandinavian Journal of Psychology,* **31**, 291–301.

[22] Ohme, R., Reykowska, D., Wiener, D., Choromanska, A. (2009). Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. *Journal of Neuroscience, Psychology, and Economics*, **2** (1), 21–31.

[23] Appel, V., Weinstein, S., & Weinstein, C. (1979). Brain activity and recall of TV advertising. *Journal of Advertising Research* , **19**, 7–15.

The Influence of Arousal on Advertising Effectiveness

D. Belanche[1], C. Flavián[2], A. Pérez-Rueda[3]

Departamento de Dirección de Marketing e Investigación de Mercados, Universidad de Zaragoza, Zaragoza, Spain.
[1] belan@unizar.es, [2] cflavian@unizar.es, [3] alfredo.perez.rueda@gmail.com

Abstract

Professionals of advertising have to create ads which get the consumers' attention and increase advertising effectiveness. This research explores the arousal concept and measurement on video ads and analyzes its relationship with ad attitude, brand recall and purchase probability. Specifically, our work in progress studies how the inclusion of high arousal sequences in ad design may affect these variables. Neuroscience and survey techniques were combined in the research methodology. Preliminary results of the study confirm that a high arousal design increases ad attitude, brand recall and purchase probability compared to a neutral arousal design. Results also suggest interesting implications for ad design and further research lines to advance on the study of advertising arousal and loyalty related variables.

Keywords: arousal, neuroscientific techniques, advertising, ad attitude, brand recall, purchase probability

Introduction

Companies invest a great amount of their budgets to deliver their messages effectively. Advertising (e.g. TV commercials) is one of the most common instruments used in marketing to build brand equity and to lead consumer to purchase. In 2013, global advertising investment reached 505,000 million (US$) which represent a 3.5% increase compared to 2012 [1]. However, people are exposed to a continuous information overload; companies are aware of this bombard with ads and explore new commercial strategies. For example, modifying advertising designs to get their audiences attention and increase communication effectiveness. Nowadays, one of the biggest challenges that advertisers have to face is to know what messages are relevant for consumers and how to stimulate consumers' mental process of advertising [2].

Our experimental design study, based on neurophysiological and survey measurement methods, aims to better understand the mechanisms by which different commercial designs in terms of arousal stimulus affect advertising effectiveness (ad attitude, brand recall and purchase probability). Our research in progress also contributes to expand the current trend to analyze individual's physiological reactions to advertising stimuli and combining it with traditional methods [3].

Literature review

Arousal: theoretical background

Arousal is a variable directly related to cognitive and affective processes, which represents the body activation level [4] as a human capacity to react to external stimuli [5]. Arousal is thus a physiological and psychological state of alertness [6]. As a fundamental dimension in the study of emotions, arousal has been related to simple processes such as awareness and attention [7], but also to more complex tasks such as information retention and attitude formation [8]. Previous literature suggests that arousal affects ad effectiveness, brand desirability [9], and memory decoding [10].

However, highly arousing contexts could distract individuals from ad processing, making recall more difficult for the focal material [11], reducing the ability to identify attributes presented in the ad, and increasing the difficulties to encode ad content [6]. Our study aims to shed some light on this theoretical debate by carrying out an empirical study linking arousal to classic variables of advertising effectiveness.

Measurement of arousal

Arousal is difficult to analyze and it has been commonly measured by self-report techniques. In contrast, some authors claim for a more accurate measurement using also physiological responses [12]. Combining different neuroscience techniques is advantageous for most marketing research fields, such as the emotional reaction measurement to advertising [3]. Functional magnetic resonance (fMRI) [13], electroencelography (EEG) [14], Electromiogram (EMG) [15], and skin conductivity [16] are surely the most relevant of neurophysiological techniques applied nowadays to better understand consumer behavior.

Hypotheses formulation

Memory, affect, and persuasion are the three basic cognitive motivation fundaments in consumer information processing [17]. Getting the viewers' attention is one of the first steps to process ad information and arousal stimulus helps to get it [10] as an impulse demanding cognitive operations [18].

Ad attitude (AAd) is probably one of the most analyzed construct in marketing research [19]. AAd is defined as ''a learned predisposition to respond in a consistently favorable or unfavorable manner to advertising in general'' [20, p. 53] and has cognitive and affective foundations [21]. AAd is a useful predictor of behavioral intentions, for this reason, a critical variable for advertising designers [22]. A cue stimulating emotional response may result in a favorable AAd, increasing the involvement and attention to the ad [21].

When consumers observe a specific exciting video sequence, they react unconsciously (high-arousal) by activating attention to achieve a more accurate memorization [10]. The stimulus acquires significance through an association process to remember other ad elements (e.g. the brand name). Accordingly, we propose that brand recall is higher for ads with a high arousal design compared with ads without them.

Additionally, often, the ultimate ad goal is to cause the greatest impact on future purchase behavior after ad exposure. Since high arousal stimuli involve consumers' reactions such as changes or emotional states activation [23], this positive stimulation may play an important role in the purchase probability too. All in all, the research hypotheses are:

Hypotheses 1-3: Compared to a neutral arousal video design, a high arousal video design has a positive influence on ad attitude (H1), brand recall (H2), and purchase probability (H3).

Method

Pretest: arousal measurement and ad selection

A pretest with a sample of 42 individuals was carried out with the aim of selecting the video ads that will be used in the experiment. The instrumentation used (electroencephalograph, skin conductance sensor and other bio-sensors) and the video presentation was synchronized within the common amplifier and the software. The participants were asked to sit comfortably and watch 24 videos of different nature. The videos were presented randomly within 3 blocks of 8 videos, with an inter-video time of 10 seconds and inter-block time of 3 minutes. This method provided an aggregated arousal measurement for each video, obtained by the arousal superposition of each subject. The video with a more variation in arousal in terms of aggregated activations was selected. The data were collected and analyzed by the company BitBrain Technologies®, and then given for the present study.

Experimental Design

Participants of the study were 73 graduate and postgraduate students of a Spanish University who were invited to participate in a marketing research and randomly assigned to each of the two conditions: with high arousal peak design (N=37), neutral design (N=36). Taking as a basis the pretest video, the two experimental video scenarios of 30 seconds were created. The first manipulated condition presented a video with a high arousal peak and 8 seconds later the brand name ("Adventure cam") during 3 seconds. The second scenario did not have any high arousal peak (neutral condition) and presented the same invented brand name at the same second than the first scenario. In all the cases, the manipulated spot was inserted in a sequence between two neutral ads (in terms of arousal) of 45 seconds each.

Measurement

AAd was measured by three items using 7-point Likert scales. Specifically, respondents had to indicate their level of agreement with the following statements: "I like the advert", "The ad get my attention" and "The ad is interesting to me" (Cronbach α =.84). Brand recall was measured by a direct question asking the name of the announced brand and coded by two external judges. Purchase probability was measured by a 7-point scale (from very low to very high probability).

Results

Participants responses analysis confirmed our hypotheses 1 and 3. In support of H1, participants in the high arousal condition presented significant higher levels of AAd on average (M=5.44; s.d.=.98) than those in the neutral condition (M=4.69; s.d.=1.30), t (70)=2.79, p=.001).

On the other hand, results reject H2. Brand recall is higher for participants in the high arousal condition compared to the neutral condition, but this difference is not significant.

Finally, in support of hypothesis 3, also the difference of purchase probability is significantly higher in the high arousal condition (M=3.08; s.d.=1.30) than in the neutral ad design condition (M=2.08; s.d.=1.16), t (70)=. 3.46, p=.01).

Conclusions and discussion

Our work in progress proposes that different high/neutral video advertising designs affects advertising effectiveness. Results lead us to the following conclusions:

First, AAd is higher in a high arousal design condition. This result suggests that consumers prefer, enjoy and get involve with high arousal advertising compared to a non-stimulating ad. Second, brand recall is marginally higher for video ads presenting a high arousal design. To certain extent, a high arousal stimulus helps get viewers' attention and memorize brand name. Thus, a neutral arousal design may involve attention activation lack and memory processes. This finding agrees with those in previous studies linking arousal, attention, and recall [9]. Finally, the study results reveal a significant difference in purchase probability between the video ads presenting high arousal design versus neutral design. Thus, this finding suggests that the high arousal design captures not only the viewers' interest but also increases consumer attraction toward the product as indicated by a higher purchase probability.

In sum, people are exposed to a continuous information overload; so much information generates consumer indifference. Our research supports that a high arousal sequence in video advertising helps to wake up the conscious and subconscious mind, getting the spectator attention and facilitating message processing.

Implications and further research lines

These preliminary findings of our research suggest management implications specifically remarkable for advertising design. Initially, our studies supports that a high arousal scene increase advertising effectiveness. Our conclusions that a higher AAd, and purchase probability are achieved with a high arousal stimulus design suggest that advertisers should include a high arousal stimulus in their ad design for a more effective communication. This finding relies on the argument that a high arousal stimulus favors a consumer positive experience and increase viewers' attention, and therefore, activates their information processing for a better brand memorization and product likeability.

Our work in progress also presents several limitations that advocate for further analyses, and interesting research lines. In this sense, one of our research limitations is that our research does not directly measure attention. Although measuring attention is not an easy task [12], a future research line could be devoted to clarify the attention role on studies linking arousal to advertising effectiveness.

Similarly, our research describes and operationalizes arousal as a general psychophysiological activation not directly related to any specific emotion. Further research lines should specifically test our proposed relationship under different emotions, and study a potential variation of arousal effects on advertising effectiveness depending on emotional states such as positive or negative valenced emotions. Our research studies could be complemented by additional measurements, different scenario designs, and logo placements [24].

References

[1] Zenith Optimedia. (2013, December 5). Advertising Expenditure Forecasts September 2013. In http://www.zenithoptimedia.com/wp-content/uploads/2013/09/Adspend-forecasts-September-2013-executive-summary.pdf

[2] Ohme, R., Reykowska, D., Wiener, D., Choromanska, A. (2009). Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. *Journal of Neuroscience, Psychology, and Economics, **2**(1)*, 21.

[3] Ariely, D., & Berns, G. S. (2010). Neuromarketing: the hope and hype of neuroimaging in business. *Nature Reviews Neuroscience, **11**(4)*, 284-292.

[4] Gould, D., Krane, V. (1992). The arousal-athletic performance relationship: Current status and future directions. In T. Horn (Eds), *Advances in Sport Psychology* (pp.119-142). Champaing IL: Human Kinetics.

[5] Benedek, M., Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, **47**(4)*, 647-658.

[6] Shapiro, S., MacInnis, D. J. (2002). Understanding program-induced mood effects: Decoupling arousal from valence. *Journal of Advertising, **31**(4)*, 15-26.

[7] Boucsein, W. (1992). *Electrodermal activity*. (2º ed.). New York: Plenum University Press.

[8] Holbrook, M. B., Hirschman, E. C. (1982). The experiential aspects of consumption: consumer fantasies, feelings, and fun. *Journal of consumer research, **9**(2)*, 132-140.

[9] Aaker, D. A., Stayman, D. M., Hagerty, M. R. (1986). Warmth in advertising: Measurement, impact, and sequence effects. *Journal of Consumer Research, **12**(4)*, 365-381.

[10] Jeong, E. J., Biocca, F. A. (2012). Are there optimal levels of arousal to memory? Effects of arousal, centrality, and familiarity on brand memory in video games. *Computers in human behavior, **28**(2)*, 285-291.

[11] Singh, S. N., Hitchon, J. C. (1989). The intensifying effects of exciting television programs on the reception of subsequent commercials. *Psychology & Marketing*, ***6***(1), 1-31.

[12] Öhman, A., Birbaumer, N. (1993). Psychophysiological and cognitive-clinical perspectives on emotion: Introduction and overview. In N. Birbaumer & A. Öhman (Eds.), *The organization of emotion: cognitive, clinical and psychophysiological perspectives.*Göttingen: Hogrefe.

[13] Plassmann, H., Kenning, P., Deppe, M., Kugel, H., Schwindt, W. (2008).How choice ambiguity modulates activity in brain areas representing brand preference: Evidence from consumer neuroscience. *Journal of Consumer Behaviour*, ***7****(4–5)*, 360–367.

[14] Bertin, M., Tokumi, R., Yasumatsu, K., Kobayashi, M., Inoue, A. (2014). Application of EEG to TV Commercial Evaluation. In *Biomedical Informatics and Technology*, 277-282.

[15] Hazlett, R. L., Hazlett, S. Y. (1999). Emotional response to television commercials: Facial EMG vs. self-report. *Journal of Advertising Research,* ***39****(1)*, 7-24.

[16] Sundar, S. S., Kalyanaraman, S. (2004). Arousal, memory, and impression-formation effects of animation speed in web advertising. *Journal of Advertising,* ***33****(1)*, 7-17.

[17] Johar, G. V., Maheswaran, D., Peracchio, L. A. (2006). Mapping the frontiers: Theoretical advances in consumer research on memory, affect, and persuasion. *Journal of Consumer Research,* ***33****(1)*, 139-149.

[18] Ohme, R., Matukin, M., Szczurko, T. (2010). Neurophysiology uncovers secrets of TV commercials. *der markt*, ***49****(3-4)*, 133-142.

[19] Bergkvist, L., Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research,* ***44****(2)*, 175-184.

[20] Lutz, R. J. (1985). Affective and cognitive antecedents of ad attitude: A conceptual framework. In *Psychological processes and advertising effects theory*, *research*, *and application* (pp. 45–63). Hillsdale, NJ: Lawrence Erlbaum Associates.

[21] Shimp, T. A. (1981). Attitude toward the ad as a mediator of consumer brand choice. *Journal of Advertising,* ***10****(2)*, 9-48.

[22] Mitchell, A. A., Olson, J. C. (1981). Are product attribute beliefs the only mediator of advertising effects on brand attitude? *Journal of Marketing Research*, ***18***, 318-332.

[23] Cheng, F. F., Wu, C. S., Yen, D. C. (2009). The effect of online store atmosphere on consumer's emotional responses–an experimental study of music and colour. *Behaviour & Information Technology,* ***28****(4)*, 323-334.

[24] Pieters, R., Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing,* ***16****(1)*, 1-16.

The impact on emotions on recall: an empirical study on social ads

A. Missaglia[1,2], M. Mauri[2], A. Oppo[2] and V. Russo[1,2]

[1]Istituto di Consumi, Comportamenti e Comunicazione, Libera Università di Lingue e Comunicazione IULM di Milano, Milano, Italia
vincenzo.russo@iulm.it
[2]Behaviour & Brain Lab, IULM University, Milan; Italy
brainlab@iulm.it

Introduction

On the basis of the scientific evidence that emotions dominate the cognitive and behavioral processes, emotions need to be considered as crucial in advertising [1]. In particular, an emotional response to an advertisement is able to influence several aspects including the attitude towards advertising and brand [2,3,4,5], attention [6,7] and finally the memory of the message and therefore its effectiveness [1,6,7,8,9,10].

Since 1879, the year Wilhelm Wundt founded the first psychology laboratory in Leipzig, many researcher focused their attention on the construct of emotion. Important authors, as William James, then proposed theories on emotions and tried to operationalize this construct. To date, the issue of emotions remains, predominantly, an issue of measurement.

Traditional techniques of investigation, categorized into those that can be described as "self-report measure”, are not able to effectively test the ads that use emotions to achieve their goals because they are too rational and verbal [11] and are distorted by cognitive processes ("cognitive bias"). Scientific research has shown that people are not aware that a lot of their daily actions are automatic and unconscious [12, 13]. Research has also shown how emotions are able to influence behavior without the person being aware of it [14]. For these reasons, methods based on subjective emotional perceptions are not always able to accurately capture the emotional state of the person. In contrast, the non-verbal measurements (or autonomic measurements), based on the registration of neurophysiological parameters, give a more accurate and reliable output because not mediated by cognitive processes that are activated during the traditional survey techniques [15].

Studies have documented the wide array of emotions that can be evoked by ads [16], examined the relationship between attitudes and consumers' responses to emotional ads [17], and investigated individual difference or situational variables that often moderate consumers' responses to emotional experiences induced by ads. However, consumer research has been largely silent about properties of emotions beyond their valence that may underlie and differentiate them. This seems to be due to the widely held assumption that the valence of an emotion featured in an ad (i.e. its positivity or negativity) is the primary predictor of a consumer's attitudinal response to the ad. Thus, if an ad depicts any of several emotions that all share a common valence (e.g., negative feelings like fear, anxiety, or guilt), people's attitudes toward the ad will simply reflect that valence (e.g., negative).

However, storytelling with emotional (positive or negative) appeals is viewed as one of the most effective strategies to encourage pro-social behavior [18].

Social campaigns often evoke negative emotions, as fear. This is a shock tactic to raise awareness and challenge preexisting attitudes toward relevant social issue as violence on women, obesity, smoking and alcohol abuse [19,20,21].

Literature has shown that there are conflicting results concerning studies that compared negative versus positive storytelling. Some authors found that the use of fear and shock in marketing campaigns has a positive effect on raising awareness about the consequences of smoking [22], the social costs of binge drinking [23] and the deleterious effects of drug abuse [24], however others have reached the opposite conclusions [25, 26] explaining this data as a “boomerang effect” described in the theory of psychological reactance [27, 28].

Aim

Given this theoretical background, we want to evaluate which kind of communication strategy is the most effective in promoting Female Genital Mutilation (FGM) awareness and we want to analyze which kind of measure (psychophysiological or self-report) predict a better recall. The present experiment was designed to assess whether arousal, assessed both with explicit measures and with psychophysiological measure was able to predict a better recall after for month.

The aim of this study was to assess the 4 months recall accuracy of two social advertising videos that used two different kind of communication strategy: one based on a non-violent code and one based on a violent code. We hypothesized that the recall was better for the video that evokes a greater arousal. Furthermore, we address if a very high arousal might interfere in recall and which is the best communicative strategy.

In particular, we want to analyse 3 different aims:

1. Assess the 4 months recall of two social advertising videos.
2. Evaluate which was the Soc Ad preferred by the participants to promote FGM awareness
3. Analyse psychophysiological and self-reported measures associated with the recall and with the video’s preference.

Design/Methodology

Forty women (age range 21-28 Mean=23.38, SD=1,97) were voluntarily recruited from the student body of IULM University of Milan (Italy). All participants signed up an informed consent before beginning the experimental procedure and were informed about the goals of the study, procedures, cautions and ethical issues for the participation to the study. Participants were randomly assigned to one of two conditions. Eighteen participants (45.0%) were first exposed at the Social Ad1 and Twenty-two participants (55.0%) were first exposed at the Social Ad2. We controlled for the effectively randomization using a chi-square experimental design (chi2=0.64; p=0.42).

Participants viewed 2 minutes-videos on this topic. The first social advertising spot (Social Ad 1) was based on a non-violent code, the second social advertising spot (Social Ad 2) was based on a violent code. The ads were selected from a range of options by an expert panel (N=10) who independently assessed as non-violent the Soc Ad 1 and as violent the Soc Ad 2. Furthermore the 100% of the experts stated that the two social advertising spots evoked different emotions, namely, sadness for the Soc Ad 1 and anger for the Soc Ad 2.

For monitoring the responses of the autonomic nervous systems, physiological signals of Blood Volume Pulse, Galvanic Skin Response (SC) and Electromyography (EMG) were monitored. Before starting the experiment every subject was exposed to 3 minutes of a black screen whit a white cross. We calculated the differences between the mean value of baseline condition and the mean value of experimental exposure.

In order to assess longer term recall about 4 month after the video exposure a phone interview was conducted by a psychotherapist**.** The interview was conducted in a structured way using mostly open-ended questions.

Results and discussion

Using a chi-square to assess the distribution of the emotion that the participants reported for both Social Ads, we observed that they report in a greater percentage to feel sadness during the Social Ad1 (chi-square=37.92, p<0.001) and anger during the Social Ad 2 (chi-square=11.64, p=0.02).

Looking at the Skin Conductance performance during the video exposures, we observed that it was significantly higher on average during the Social Ad2 (F=3.88, p=0.05). Overall, we can state that the level of arousal was significantly greater during the view of the Social Ad2 in comparison to Ad1. Skin Conductance has been showed as a good indicator of the level of arousal correlating with the level of anxiety elicited by the stimuli [29].

Moreover we found that both psychophysiological measure and self-report measure was congruent. In fact, participants had both a greater state anxiety (STAI-X) and a higher level of Skin Conductance in viewing the social ad that used a violent code communication. Regarding the EMG signal, we found that it was similar in participants during the viewing of the two Social Ads. Our results on psychophysiological signals (SC and EMG) are in line with the Lang model (1995), in fact the two spots are design to have the same valence (aversive) but different arousal level to test only the effectiveness of anxiety/stress (arousal) on recall. Furthermore, our results are consistent with other studies conducted in the field of advertising that identified in the SC the best indicator for arousal that is an index of an emotional reaction.

Regarding our first aim, that was identifying which was the better strategy to increase awareness on FGM, we found that the Social Ad that used a violent code was significantly better recalled at 4 month after video exposure in terms of accuracy and number of details reported. Our data are in line with the literature that reports how, in social advertising, a shock communication strategy aimed to raise awareness on relevant social topics might represent an efficient way to use.

Furthermore, after the exposure to each spot, subjects were instructed to click on the kind of emotion they felt during the vision (they were instructed to choose the emotion amongst a list presented on the pc screen, clicking on one of the following options: anger, joy, sadness, fear, surprise, disgust. See Figure1 and 2). We found that participants who reported to feel anger (mean RT=9.38 sec; SD=4.96 sec) during the exposure to Soc Ad2 were faster (t Student test: p=0.035) in clicking with the mouse on the label “anger” in comparison to time reactions of participants who reported to feel other type of emotions (mean RT=15,56 sec; SD=10,13 sec).

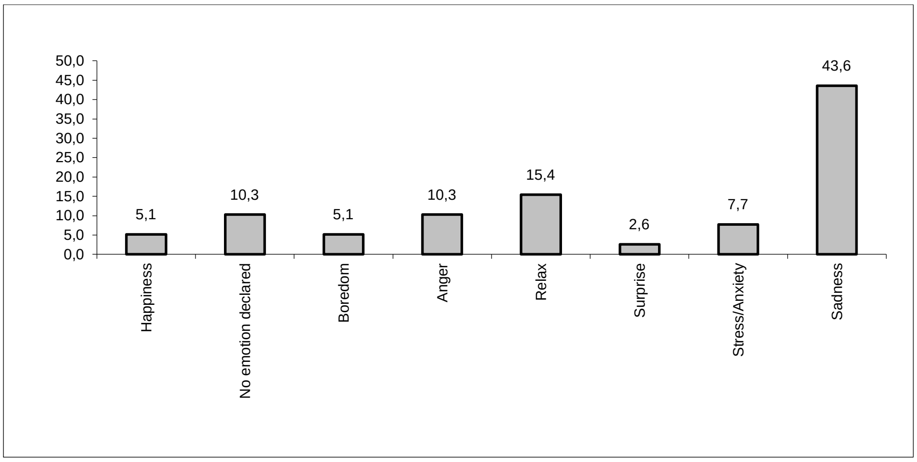


Figure 1. Self-report scores (average across all subjects) after Social Ad1 (soft communication strategy) exposure: the highest score is for “Sadness”, with a total of 43,6 % of all subjects reporting to feel “sadness” after being exposed to the Social Ad1. The second highest score is “relax” (15,4%). Nobody reported “Disgust”.

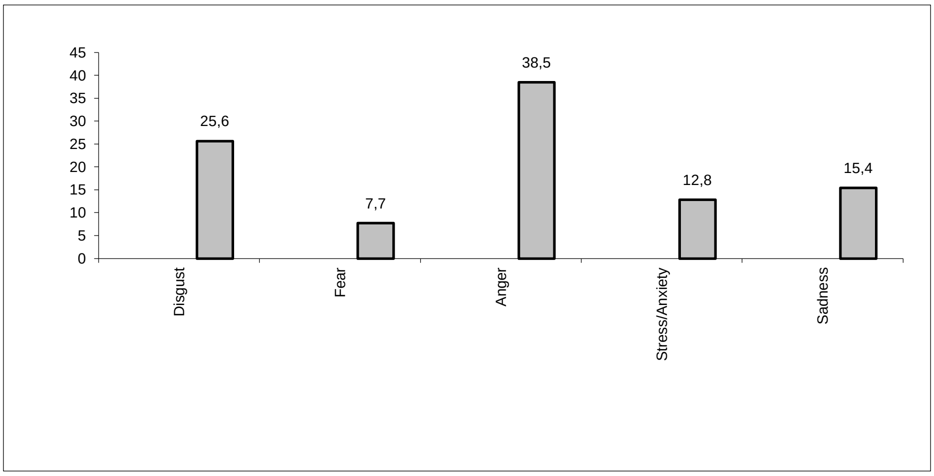


Figure 2. Self-report scores (average across all subjects) after Social Ad2 (violent communication strategy) exposure: the highest score is for "Anger", with a total of 38,5 % of all subjects reporting to feel "Anger" after the exposure to the spot. It is worthy to note that the second highest score is "Disgust", with 25,6 %.

It could be hypothesized that the congruence between explicit and psychophysiological data might be explained with the topic of the social advertising spots used in our research. Female Genital Mutilation, in western countries, is a practice largely condemned (Morrone A., 2002), for this reason the participants, in line with social desirability theory (Edwards, 1957), are free to declare their opinion and feeling about this topic. However, it could be argued that in other sociocultural contexts or with a sample of foreign women we could find incongruent results between explicit and implicit measures. Thus, it could be very useful assess this phenomenon both with traditional assessment and with measures used in neuromarketing evaluations because foreign women may not feel free to express their feelings about a practice such as FGM, which is widely prevalent in their culture.

In this study, the 30% of participants declared that they would choose the non-violent social ad for promoting an efficient social campaign and this percentage almost doubled (64.3%) if we consider only participants who have reported to feel anger during the viewing of the Social Ad that used a violent-communication strategy. Taking advantage of this finding, we found that the likelihood of choosing a Social Ad that used a non-violent strategy was about ten-fold higher in those participants who felt anger during the viewing of the Social Ad that used a violent-communication strategy. Furthermore, we wonder if, regarding the issue of Female Genital Mutilation, the best way to enhance the awareness of this delicate topic is a shock tactic that, as we have seen, is associated with a better recall, but at the same time, we have to consider that we could meet the risk of evoke a strong emotion as anger that might contribute to create distance between the Italian and the foreign population.

Our findings suggest some interesting issues in the field of social communication. First of all, in this work the combination of self-reports and psychophysiological measures is explored, as the research study is related to a topic affected by cultural and sexual issues. For this reason, self-reports and biological measures might reveal different trends, that was not the case for this study as shown in previous research work [30]. It is fundamental that Social cause advertising researchers stay in touch with developments in emotion research from rapidly developing fields like experimental psychology and neuroscience. Neuro-physiological measures may be a useful and objective supplement to subjective, declarative data. When combined, these two forms of modality may enable marketers to represent both conscious and subconscious consumer reactions to persuasive advertising [12,31,32].

References

[1] Hazlett, R., Hazlett, L., Yassky, S. (1999). Emotional Response to Television Commercials: Facial EMG vs. Self-Report. *Journal of Advertising Research,* ***39****(2),* 7-23.

[2] Batra, R., Ray, M.L. (1986). Affective Responses Mediating Acceptance of Advertising. *Journal of Consumer Research,* ***13****(2),* 234-50.

[3] Edell, J.A., Burke, M.C. (1987). The Power of Feelings in Understanding Advertising Effects. *Journal of Consumer Research,* ***14****(3),* 421-33.

[4] Du Plessis, E. (2005). *The Advertised Mind.* London, UK: MillwardBrown.

[5] Hall, B.R. (2002). A New Model for Measuring Advertising Effectiveness. *Journal of Advertising Research,* ***42****(2),* 23-31.

[6] Medina, J. (2010). *Il cervello. Istruzioni per l'uso*. Bollati Boringhieri

[7] Stayman, D.M., Batra, R. (1991). Encoding and Retrieval of Ad Affect in Memory. *Journal of Marketing Research,* ***28(****2),* 232-39

[8] Li-Wei, M., Scholler G. (2009). Emotions, attitudes and memorability associated with tv commercials. *Journal of Targeting, Measurement and Analysis for Marketing,* ***17***, 55-63.

[9] English, B.C. (1990). Consumer emotional reactions to television advertising and their effects on message recall. In S. Agres, J.A. Edell, and T.M. Dubitsky (Eds.), *Emotion in advertising: Theoretical and practical explorations*. Westport, CT: Quorum Books.

[10] Page, T.J., Thorsom, E., Heide, M.P. (1990). The Memory Impact of Commercials Varying in Emotional Appeal and Product Involvement. In S. Agres, J.A. Edell, and T.M. Dubitsky (Eds.), *Emotion in advertising: Theoretical and practical explorations*. Westport, CT: Quorum Books.

[11] Haley, R.I., Staffaroni, J., Fox, A. (1994). The Missing Measures of Copy Testing. *Journal of Advertising Research,* ***34***, 46-41.

[12] Zaltman, G. (2003). *How Customers Think*. Boston, MA: Harvard Business School Press.

[13] Chartrand, T.L. (2005). The Role of Conscious Awareness in Consumer Behavior. *Journal of Consumer Psychology,* ***15(****3),* 20.

[14] Winkielman P., Berridge, K.C., Wilbarger, J.L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin,* ***31****(1),* 121-35.

[15] Poels, K., DeWitte, S. (2006). How to Capture the Heart? Reviewing 20 Years of Emotion Measurement in Advertising. *Journal of Advertising Research,* ***46(****1),* 18-37.

[16] Holbrook, M.B., Batra, R. (1987). Assessing the Role of Emotions ad mediators of Consumer Responses to Advertising. *Journal of Consumer Research,* ***14***, 404-420.

[17] Burke, M.C., Edell, J.A. (1989). The impact of feelings on ad-based affect and cognition. *Journal of marketing research,* ***26***, 69–83.

[18] Merchant, A.J., Ford, B., Sargeant, A. (2010). Charitable organizations' storytelling influence on donors' emotions and intentions. *Journal of Business Research,* ***63(****7),* 754-762.

[19] Henthorne, T.L., LaTour, M.S., Nataraajan, R. (1993). Fear appeals in print advertising: An analysis of arousal and ad response. *Journal of Advertising,* ***22****(2),* 59– 69.

[20] LaTour, M., Herbert, S., Rotfeld, J. (1997). There are threats and (maybe) fear-caused arousal: theory and confusions of appeals to fear and fear arousal itself. *Journal of Advertising,* ***26****(3),* 45-60.

[21] Witte, K., Allen, M. (2000). A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Campaigns. *Health Education Behavior,* ***27****(5),* 591-615.

[22] Wakefield, M., Flay, B., Nichter, M., Giovino, G. (2003). Role of media in influencing trajectories of youth smoking. *Addiction,* ***98****(1),* 79-103.

[23] Pilling V.K., Brannon, L.A. (2007). Assessing college students' attitudes toward responsible drinking messages to identify promising binge drinking strategies. *Health Communication,* ***22****(3),* 265-276.

[24] Schmeling, D.G., Wotring, C.E. (1980). Making anti-drug abuse advertising work. *Journal of Advertising Research,* ***20****(3),* 33-37.

[25] Capella, M. L., C. R. Taylor and C. Webster. (2008). Cigarette Advertising Bans' Effect on Consumption: A Meta-Analysis. *Journal of Advertising,* ***37****(2),* 7-18.

[26] Pechmann, C., Ratneshwar, S. (1994). The Effects of Antismoking and Cigarette Advertising on Young Adolescents' Perceptions of Peers Who Smoke. *Journal of Consumer Research,* ***21***, 236-251.

[27] Brehm, J.W. (1966). *A theory of psychological reactance*. New York: Academic Press.

[28] Brehm, S.S., Brehm, J.W. (1981). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.

[29] Mauri, M., Magagnin, V., Cipresso, P., Mainardi, L., Brown, E.N., Cerutti, S., Villamira, M., Barbieri, R. (2010). Psychophysiological signals associated with affective states. *Proceedings of 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires, pp. 3563–3566.

[30] Falk E.B., Berkman, E.T., Lieberman, M.D. (2012). From Neural Responses to Population Behavior: Neural Focus Group Predicts Population-Level Media Effects. *Psychological Science,* ***23****(5)*, 439–445.

[31] Damasio002C A. (1995). *L'errore di Cartesio. Emozione, ragione e cervello umano*. Adelphi; ed. orig.: “Descartes' Error: Emotion, Reason, and the Human Brain”, Putnam, 1994.

[32] LeDoux, J.E. (1994). Emotion, Memory and the Brain. *Scientific American,* ***270****(6)*, 50-57.

Game Over!
On measurement and optimization of presence

Egon L. van den Broek and Remco Kuijper

Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, the Netherlands
vandenbroek@acm.org

Abstract

With reality and ICT blending, we live in a Computer Aided Reality (CAR). Nevertheless, we still have the feeling that we can distinguish between reality and artificial realities. This notion yields the question: When do humans experience reality as computer aided (or artificial)? This article contributes in answering this question by i) adopting the construct presence and discuss its value to assess users' experienced reality; ii) providing suggestions on how to measure presence via biosensors; and iii) presenting a low fidelity research setup to assess users' presence in CAR.

Introduction

The influence of ICT on our society shows a constant, monotonic increase. It is hard to calculate its impact, as it changes and interacts with virtually all of society's aspects. Nevertheless, its influence is undeniably there. One of its manifestations is virtual worlds and, in particular, Virtual Reality (VR), which have been promising for over 20 years. Regrettably, up to the start of this decennium, due to their technical complexity and cost, VR has never made it to the general public [1]. However, throughout the last years several low cost technologies have been introduced that can help to alter reality: Microsoft's Kinect provides a low cost object tracking device [2]; Oculus Rift provides a Virtual Reality headset (or Head Mounted Display, HMD) for 3D gaming[1]; and Jaguar's Virtual Windscreen[2] provides advanced augmented reality for their cars. Moreover, low cost high-level programming languages such as 3D Studio Max (Autodesk, Inc.) and Quest3D (Act-3D B.V.) enable the generation of virtual worlds [3].

Recent ICT progress yields the question: *When do humans experience reality as computer aided (or artificial)?* Measuring behavior in reality is already well studied, mainly in social and behavioral sciences. However, behavior is less studied in Computer Aided Reality (CAR) [4,5] and, consequently, a comparison cannot be made in a straightforward manner. Most often the construct presence is used to indicate user's experience in CAR. Therefore, we want to develop a measurement technique to determine the users' presence, in any reality, including CAR. This would provide us with a measurement tool to determine the distance in users' experienced presence between a CAR and everyday life's reality.

What is it that we call reality?

In Book VII of *The Republic*, dated 360 B.C., Plato describes his allegory of the cave. He questions our reality, which could as easily be an illusion. This touches upon the foundations of philosophy and physics; but also on what we now denote as CAR. Both traditional branches of science question axioms that are often taken for granted. With ICT's progress, one can question what axioms can be maintained in a CAR, where all that seems to matter is what we experience. So, with the current technology push, Plato's work is perhaps more timely than ever. For the interested reader, see also Kant's work on analogies of experience [6].

Here, we will illustrate basic notions such as mentioned above and illustrate their use in daily practice. To limit the scope of the basic notions and bring it to engineering practice, we define CAR: any reality that is partially

[1] URL: http://www.oculusvr.com/ [Last accessed on July 17, 2014]
[2] URL: http://www.digitaltrends.com/cars/jaguars-virtual-windscreen-turns-real-life-into-a-video-game/ [Last accessed on July 17, 2014]

generated, designed, or altered via information, computing, and/or communication technology. CAR can interact either passively (e.g., as in a movie or virtual tour) or actively (e.g., biofeedback or augmented reality). Either way, it is perceived by at least one of our senses (not necessarily vision). It includes VR, virtual worlds, and games but also mixed reality, augmented reality, and reality itself (in particular, as envisioned by Ambient Intelligence, AmI).

Presence (and immersion)

The term presence is often interchangeably used with the term immersion. According to Slater et al. [7] presence and immersion describe two different concepts. Immersion refers to a quantifiable set of technological properties, whereas presence refers to a user's subjective state of consciousness. With a high sense of presence, judgments about a product's design can be considered more valid and reliable. Whether a user feels present in a CAR depends on a large number of factors. Presence is affected by the technological, immersive aspects of the CAR and the personal characteristics of the individual [7; cf. 8,10].

Slater et al. [7] stated that the immersive character of CAR (e.g., VR and games) is determined by the:

- number of sensory systems (in particular vision, sound, and touch),
- extent that information is provided from any direction,
- extent that external noise is excluded,
- correspondence between the user's behavior and the system's feedback, and
- degree of sensory richness - or realism.

The variables associated with these technological characteristics that define the level of immersion include the field of view, perceptual realism, multimodality, free control of actions, and digital representation of the body. Variables associated with the individual that make up the user's experienced presence include motivation, ability, willingness, flow, and prior experience, amongst others [7,9]. The former set of variables are defined by the system's technical specification. The latter set of variables is harder to pinpoint.

Between realities: Closing the gap

There is no single, accepted paradigm for the assessment of presence [4]. Often the approach is twofold: subjective measures and objective corroborative measures (e.g., posture and physiological responses). Various questionnaires have been introduced and nearly all showed to be useful. In contrast, corroborative measures have been used little and, so far, have not shown to be reliable. Consequently, mainly questionnaires are used [4,5]. However, these questionnaires only suffice for offline, post-hoc research or, alternatively, users' sense of presence is interrupted to ask users to complete them, which should be prevented.

There are two traditional approaches in tackling the behavioral measurement problem of capturing user's presence: i) data driven: continuous autonomous feedback mechanisms and ii) knowledge driven: a nomological model on presence. On the one hand, we adopt the data driven approach in our effort to capture user's sense of presence. For this goal, we can use biosensors to capture users' sense of presence as it suits a well-established continuous autonomous feedback mechanism for various domains and applications and biosensors can be embedded in various products [10,11]. On the other hand, we aim to add knowledge to the theoretical framework of presence. For this approach, we suggest to explore four theoretical frameworks:

1. Csikszentmihalyi's Flow-theory: The psychology of optimal experience [9,12];
2. Russel's arousal, valence, and dominance model [8,11];
3. Ekman's discrete, basic emotions [8,13]; and
4. Silvia's appraisal theory of interest. "This suggests a sweet spot of interest, where information that is novel and complex but still comprehensible." [12].

Each of these four theoretical frameworks is established but, simultaneously, debated [14] and includes constructs related but different from presence. Riva et al. [13] describe "a circular relation between presence and emotions: on the one side, the feeling of presence was greater in the ‘emotional’ environments; on the other side, the emotional state was influenced by the level of presence”. Baños et al. [15] reported similar conclusions, although they also reported an interaction effect between emotion and immersion on presence. However, their experimental sample sizes were small, emotion and presence were solely assessed by questionnaires, and Riva et al. [13] only tested two emotional states. The question is how the four frameworks and the affiliated empirical evidence relate to each other and how much of the recorded variance on the construct presence they can explain. To enable cross-validation of the continuous autonomous feedback mechanism, existing, validated presence questionnaires will be used.

Unveiling the level of presence: Methods and apparatus

Riva et al. [13] and Baños et al. [15] unveiled a relation between emotions and presence. However, their results seem to suggest that emotions fulfill a mediating role in user’s experienced presence. Either way, biosensors can be employed to determine emotions [8,9,10,11,13] and, consequently, can provide valuable information on experienced presence. We used the low cost, modular, wearable BITalino[3] sensor kit to measure biosignals (sample rate: 1000 Hz; 10-bit/6-bit; Class II Bluetooth v2.0 data transmission) [16]. For details on the BITalino sensor kit, see their website and [16].

Emotion is decomposed in its two main dimensions: arousal and valence [8,10,11,15]. Arousal is assessed using ElectroDermal Activity (EDA), recorded midway between the proximal phalanx and a point directly beneath the ankle to prevent motion artifacts [17]. Valence is assessed using facial ElectroMyoGraphy (fEMG) of the corrugator supercilii and zygomaticus major [11]. ElectroCardioGraphy (ECG) is recorded, using the modified Lead II placement, which can unveil information concerning both dimensions [8]. With all these sensor that record these signals, with each participant a new set of standard pre-gelled, self-adhesive, disposable electrodes was used. The placement of the electrodes for each of the signals is shown in Figure 1. Additionally, a LUX light sensor was placed on the PC’s TFT screen for the synchronization between the PC and the BITalino sensor kit.

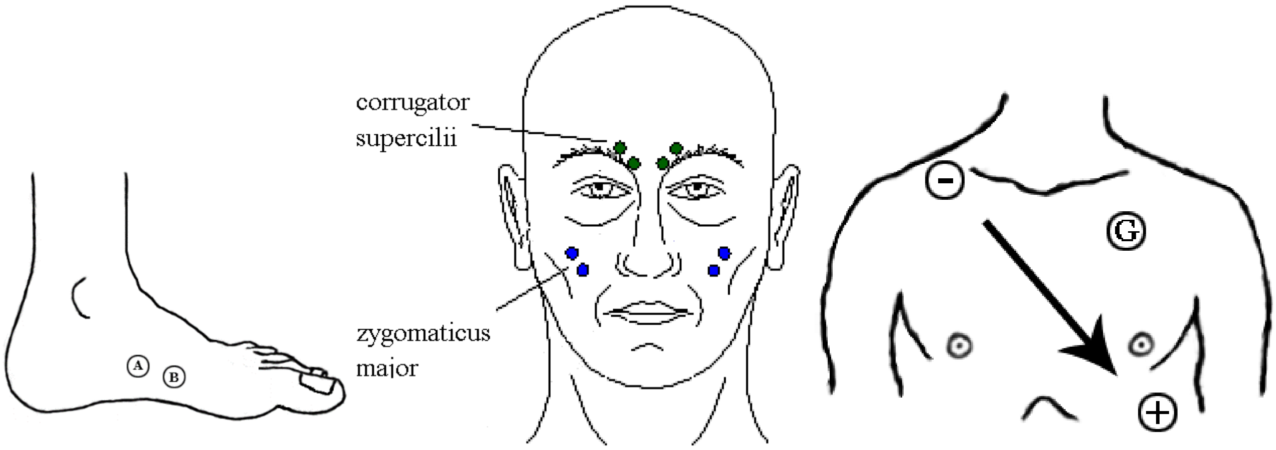


Figure 1. Positioning of the electrodes. From left to right, the electrodes for electrodermal activity on the foot (adapted from [17]), facial electromyography of two muscles (adapted from [18]), and the modified Lead II placement for electrocardiography (G denotes the ground ground electrode).

[3] URL: http://www.bitalino.com [Last accessed on July 17, 2014]

Figure 2. Left: Screenshot of the game, where a weapon bought disappeared. Right: A four-dimensional Likert scale questionnaire to assess the emotional state of the player, as was prompted after each of the three games.

Instead of an expensive full fetched and expensive VR environment (e.g., a CAVE [1]) or an obtrusive apparatus such as Oculus Rift[1], we have chosen a game as our test bed. The game Call of Duty: World at War (i.e., a first-person shooting game) [19] is used to assess users' presence when playing it. To assure an optimal immersion, frame rate was set on 30 fps with high quality graphics to maximize the visual realism [3]. In the game's zombie mode, the goal is to survive as many waves of enemy zombies as possible. With every passing wave, the number and speed of opposing zombies increases. The incremental nature of the difficulty level is in line with Csikszentmihalyi's theory of flow for the optimal experience (mostly for beginning players) [9] and, hence, is expected to be a viable candidate for measuring presence. The acceleration in difficulty was increased with 75% per wave of zombies, compared to the original settings. This limits the playing time of advanced players and, more importantly, assures users to allocate sufficient workload and attention to the game and, hence, to increase the chance on a sufficient level of presence. Additionally, the game was altered in various ways to force a, so called, break in presence (BIP) [7]. A BIP is defined as a transition in response patterns from virtual to real [7]. These adaptations included unexpected teleportation, disappearing of weaponry, and a change of type of symbols used in the game; see also Figure 2, left. The orders of execution of these adaptations were experimentally controlled. No further adaptations to the game were made. Each of the participants both played the game and, subsequently, answered four questions. five times; see also Figure 2. Hence, essentially, the participants were able to play the game without any constraints.

Conclusions

Reality is altered by ICT. In practice, we are already living in a Computer Aided Reality (CAR). Traditionally, CAR was considered to be a virtual world, a game, VR, or something related. However, with reality and ICT blending (creating AmI or the Internet of Things) and a strict separation of both becomes more and more impossible, this division seems to lose its usefulness [cf. 6,8]. This raises attention to Plato's question [cf. 6]: What is reality? However, nowadays, the more interesting question is: When is reality experienced artificially? To answer this question, we pose to use the construct presence [4,5,7] and exploit its relation to emotions [13,15]. This article contributes in answering this question by i) discussing the complexity of the constructs reality and presence; ii) provide concrete suggestions on how to measure presence; and iii) presenting a low cost, modular, wearable research setup [16] to assess user's level of presence in a gaming environment [19]. This approach can be easily extended to other CAR (e.g., augmented reality) and with the use of other (embedded) biosensors [20].

References

[1] Meng, F., Zhang, W., Yang, R. (2014). The development of a panorama manifestation virtual reality system for navigation and a usability comparison with a desktop system. *Behaviour & Information Technology, **33**(2)*, 133-143.

[2] van Teijlingen, W., van den Broek, E.L., Könemann, R., Schavemaker, J.G.M. (2012). Towards sensing behavior using the Kinect. In *Proceedings of Measuring Behaviour 2012: 8th International Conference on Methods and Techniques in Behavioural Research*, 28-31 August, Utrecht, The Netherlands. pp. 372-375. Wageningen, The Netherlands: Noldus IT.

[3] Meijer, F., Geudeke, B.L., van den Broek, E.L. (2009). Navigating through virtual environments: visual realism improves spatial cognition. *CyberPsychology and Behavior, **12**(5)*, 517-521.

[4] Norman, K.L. (2013). GEQ (Game Engagement/Experience Questionnaire): A review of two papers. *Interacting with Computers, **25**(4)*, 278-283.

[5] Behm-Morawitz, E. (2013). Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in Human Behavior, **29**(1)*, 119-128.

[6] Swabey, W.C. (1922). Kant's analogies of experience. *The Philosophical Review, **31**(1)*, 41-57.

[7] Slater, M., Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, **9**(5)*, 413-434.

[8] van den Broek, E.L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing, **17**(1)*, 53-67.

[9] van den Broek, E.L. (2012). On making engagement tangible. In *Proceedings of Measuring Behaviour 2012: 8th International Conference on Methods and Techniques in Behavioural Research*, 28-31 August, Utrecht, The Netherlands. pp. 90-93. Wageningen, The Netherlands: Noldus IT.

[10] van den Broek, E.L., Janssen, J.H., Westerink, J.H.D.M. (2015). Autonomous closed-loop biofeedback: An introduction and a melodious application. In R.A. Calvo, S.K. D'Mello, J. Gratch, and A. Kappas (Eds.), *The Oxford Handbook of Affective Computing (Section 5: Applications of Affective Computing), Chapter 35*, p. 472-482. Series: Oxford Library of Psychology. New York, NY, USA: Oxford University Press, Inc.

[11] Westerink, J.H.D.M., van den Broek, E.L., Schut, M.H., van Herk, J., Tuinenbreijer, K. (2008). Computing emotion awareness through galvanic skin response and facial electromyography. In *Probing Experience: From Academic Research to Commercial Propositions (Part II: Probing in order to feed back), Chapter 14*, p. 137–150. Dordrecht, The Netherlands: Springer Science + Business Media B.V.

[12] van der Sluis, F., van den Broek, E.L., Glassey, R.J., van Dijk, E.M.A.G., de Jong, F.M.G. (2014). When complexity becomes interesting. *Journal of the American Society for Information Science and Technology, **65**(7)*, 1478–1500.

[13] Riva, G., Mantovani, F. Capideville, C.S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychology & Behavior, **10**(1)*, 45-56.

[14] Lindquist, K.A., Siegel E.H., Quigley, K.S., Barrett, L.F. (2013). The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin, **139**(1)*, 255-263.

[15] Baños, R.M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., Rey, C. (2004). Immersion and emotion: Their impact on the sense of presence. *CyberPsychology & Behavior, **7**(6)*, 734-741.

[16] da Silva, H.P., Lourenço, A., Fred, A., Martins, R. (2014). BIT: Biosignal Igniter Toolkit. *Computer Methods and Programs in Biomedicine, **115**(1)*, 20-32.

[17] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures (2012). Publication recommendations for electrodermal measurements [Committee report]. *Psychophysiology, **49**(8)*, 1017–1034.

[18] van Boxtel, A. (2010). Facial EMG as a tool for inferring affective states. In *Proceedings of Measuring Behaviour 2010: 7th International Conference on Methods and Techniques in Behavioural Research*, 24-27 August, Eindhoven, The Netherlands. pp. 104-108. Wageningen, The Netherlands: Noldus IT.

[19] Treyarch (2008). *Call of duty: World at War* [PC]. Santa Monica: Activision Blizzard, Inc.

[20] Van den Broek, E.L., Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics, **40**(6)*, 1055–1064.

Circadian and homeostatic sleep regulation:

Measuring clock-homeostat interactions

Tom Deboer

Laboratory for Neurophysiology, Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherland
Tom.de_Boer@lumc.nl

Sleep is regulated by homeostatic and circadian processes [1]. In mammals, homeostatic sleep pressure is reflected by electroencephalogram (EEG) slow-wave activity (SWA, EEG power density between ~1 and 4 Hz) in undisturbed non-rapid eye movement (NREM) sleep. In all mammalian species investigated until now, not only the amount of sleep, but also SWA in NREM sleep increases after sleep deprivation. In several species a dose-response relationship between waking duration and subsequent SWA has been established and in humans taking an afternoon nap a predictable decrease in the nocturnal NREM sleep SWA was observed. Mathematical models predicting this phenomenon have been successfully applied in humans, rats and mice [2,3].

The circadian process is controlled by an endogenous pacemaker, located in the suprachiasmatic nucleus (SCN) of the hypothalamus [4]. This pacemaker is thought to provide the sleep homeostat with a circadian framework. The circadian clock has a molecular basis for generating electrical activity rhythms in the SCN. This electrical neuronal activity can be recorded in vivo in freely moving animals [5]. Kept in constant conditions, this activity is high during the subjective day, the part of the animals rhythm that normally falls in the light period, and low during the subjective night.

Whether homeostatic and circadian regulation of sleep work independent or interact closely has been subject of many discussions. Homeostatic responses in sleep persist after circadian rhythmicity has been abolished by SCN lesioning,[6,7] and the circadian process can be manipulated by light in the morning without changing SWA [8]. It has, therefore, long been assumed that the timing of sleep is regulated independent from the need for sleep. However, more recent data indicate that there may be a continuous interaction between sleep homeostasis and the circadian clock. [9,10].

To get as close as possible to both processes, we set out to record both signals simultaneously in vivo in rats. All experiments were performed under the approval of the Animal Experiment Committee of the Leiden University Medical Center according to the Dutch law on animal experiments. The animals were equipped with electrodes recording EEG and EMG, to determine vigilance states and SWA in NREM sleep, together with electrodes recording SCN neuronal activity. Before the experiments, the animals were kept in constant darkness for at least a week, to exclude direct influences of light on behaviour, and all signals were recorded simultaneously on the same computer. In addition, drinking behaviour was recorded to obtain an estimate of rest-activity behaviour.

From baseline recordings it became clear that SCN neuronal activity changed under influence of vigilance state changes. Activity was increased during waking and REM sleep and decreased during NREM sleep. In addition, there was a clear negative correlation between EEG SWA and SCN neuronal activity. Suppression of SWA by slow-wave deprivation resulted in increased SCN neuronal activity, whereas REM sleep deprivation resulted in decreased activity. Total sleep deprivation caused the predicted increase in SWA during subsequent NREM sleep, and simultaneously a long term suppression of SCN neuronal activity. The latter lasted as long as the increase in SWA needed to recover [11,12].

Neuronal activity in the SCN clearly depended on the vigilance state of the animals, and deeper NREM sleep was accompanied by lower SCN neuronal activity. The data suggest an interaction between sleep homeostatic mechanisms and the circadian clock in which the clock receives continuous information about the status of the homeostatic process. Our present work is concentrating on possible consequences of this interaction on the circadian system, particularly when homeostatic sleep pressure is increased due to sleep loss. Sleep deprivation seems to diminish this functioning, as it was shown that the circadian clock phase shifting response to light is attenuated in sleep deprived animals [13,14]. With our techniques we are trying to resolve the neurophysiological mechanism behind this phenomenon and to investigate whether this can be reversed by pharmacological means.

References

[1] Daan, S., Beersma, D.G.M., Borbely, A.A. (1984) Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol* **246**, R161-R178.

[2] Yasenkov, R., Deboer T. (2012) Circadian modulation of sleep in rodents. *Progress in Brain Res* **199**, 203-218.

[3] Deboer, T (in press) Behavioral and electrophysiological correlates of sleep and sleep homeostasis. *Curr Top Behav Neurosci.*

[4] Meijer, J.H., Rietveld, W.J. (1989) Neurophysiology of the suprachiasmatic circadian pacemaker in rodents. *Physiol Rev* **69**, 671-707.

[5] Meijer, J.H., Watanabe, K., Schaap, J., Albus, H., Detari, L. (1998) Light responsiveness of the suprachiasmatic nucleus: long-term multiunit and single-unit recordings in freely moving rats. *J Neurosci* **18**, 9078-9087.

[6] Mistlberger, R.E., Bergmann, B.M., Waldenar, W., Rechtschaffen, A. (1983) Recovery sleep follwing sleep deprivation in intact and suprachiasmatic nuclei lesioned rats. *Sleep* **6**, 217-233.

[7] Tobler, I., Borbely, A.A., Groos, G. (1983) The effect of sleep deprivation on sleep in rats with suprachiasmatic lesions. *Neurosci Lett* **42**, 49-54.

[8] Dijk, D.J., Beersma, D.G.M., Daan S. (1989) Bright morning light advances the human circadian system without affecting NREM sleep homeostasis. *Am J Physiol* **156**, R106-R111.

[9] Edgar, D.M., Dement, W.C., Fuller, C.A. (1993) Effect of SCN-lesions on sleep in squirrel monkeys: evidence for opponent processes in sleep-wake regulation. *J Neurosci* **13**, 1065-1079.

[10] Achermann, P., Borbely, A.A. (1994) Simulation of daytime vigilance by the additional interaction of a homeostatic and a circadian process. *Biol Cybern* **71**, 115-121.

[11] Deboer, T., Vansteensel, M.J., Detari, L., Meijer, J.H. (2003) Sleep states alter activity of suprachiasmatic nucleus neurons. *Nat Neurosci* **6**, 1086-1090.

[12] Deboer T., Detari, L., Meijer, J.H. (2007) Long term effects of sleep deprivation on the mammalian circadian pacemaker. *Sleep* **30**, 257-262.

[13] Mistlberger, R.E., Landry, G.J., Marchant, E.G. (1997) Sleep deprivation can attenuate light-induced phase shifts of circadian rhythms in hamsters. *Neurosci Lett* **238**, 5-8.

[14] Challet, E., Turek, F.W., Laute, M, Van Reeth, O. (2001) Sleep deprivation decreases phase-shift responses of circadian rhythms to light in the mouse: role of serotonergic and metabolic signals. *Brain Res* **909**, 81-91.

Cross Wavelet Analysis to Study Periodic Behaviour of Animals in Relation to Environmental Cues

Maitreyi Sur[1], Andrew K. Skidmore[1], Klaus-Michael Exo[2], Tiejun Wang[1], Bruno Ens[3] and A. G.Toxopeus[1]

[1]Department of Natural Resources, Faculty of Geo-Information Science and Earth Observation, University of Twente, PO Box 217, 7500 AE, Enschede, The Netherlands
m.sur@utwente.nl, a.k.skidmore@utwente.nl, t.wang@utwente.nl, a.g.toxopeus@utwente.nl
[2] Institute of Avian Research,"Vogelwarte Helgoland", An der Vogelwarte 21,D-26386 Wilhelmshaven, Germany
michael.exo@ifv-vogelwarte.de
[3] SOVON Dutch Centre for Field Ornithology & Netherlands, P.O. Box 59, Den Burg (Texel) 1790 AB, The Netherlands
Bruno.Ens@sovon.nl

Introduction

Individual based studies of animal movement have recently gained momentum due to extensive accessibility to telemetry data from tagged animals. A key challenge in understanding animal behaviour from movement data has been explaining the role of environmental cues. This is because animals moving through complex environments are influenced by multifarious factors, and it becomes challenging to extract the influence of each environmental control at the right scale. To understand the decision making process during animal movement triggered by organism–environment information fluxes, there is a need to link environmental data with statistics of animal movement.

In order to confront these challenges we propose a statistically robust method that extracts distinct patterns in moving animals as repetitive cyclic behaviour. In case of animals, repetitive or periodic behaviour can be expected to be a result of periodic cues from nature such as temporal oscillations in temperature, precipitation, light, wind, forage etc. as well as the animal's circadian rhythm. The periodic movement pattern associated with environmental cue is likely to have a degree of temporal correlation and also operate at multiple scales [1]. Since the period (the regular time intervals in a periodic motion) that is affected by different cues is an unknown parameter, one needs a signal processing method that automatically detects periodicity in the data. Among the different period detection techniques that can be applied to signal processing, wavelet transform can be directly applied to movement data [2] and results in exploratory analysis of the auto-correlative properties of the movement data. In this study we show the use of a novel method in animal movement studies by applying a bivariate extension of the wavelet transform, namely cross wavelet analysis, to extract periodic patterns in moving animals in relation to environmental cues. First, significant locations that can be designated as reference spots (such as nesting sites) were identified from which the periodicities can be viewed. Cross wavelet analysis between a movement parameter and an environmental parameter was then used to study repetitive movement behaviour in animals. We use empirical data of GPS tagged Lesser Black-backed gulls (Larus fuscus) to assess our method.

Material and Method

Cross wavelet analysis was applied in two phases:

1. Step length during the breeding season of a single gull (ID 41781) was tested against three environmental variables, namely temperature, precipitation and wind speed.
2. Step length of 12 gulls during the breeding season was tested against temperature to illustrate individual variations and group behaviour.

Tracking data

The tracking data of the Lesser Black-backed Gull (here after gull) used for this study was provided by the Avian Research Institute, Wilhelmshaven, Germany and SOVON Dutch Centre For Field Ornithology, Nijmegen. Each bird was equipped with an Argos-GPS solar-powered Platform Terminal Transmitter (PTT; Microwave Telemetry Inc., Columbia, MD, USA). The Argos-GPS PTTs have an accuracy of ±18 m. For this study the gull with the PTT id 41781(sampling interval 1 hour) was used to test the effect of animal movement with different environmental variables at different temporal scales. Data from twelve gulls (PTT id's 41742, 41749, 41752, 41757, 41758, 41762, 41763, 41764, 41767, 41771, 41775, and 41781) were used to demonstrate the use of cross wavelet analysis in analyzing individual variations and group pattern of gulls with temperature as the environmental variable. We selected the month of May, during the breeding season to test our methods. Animals with periodic behaviour commonly have a central location that they visit repeatedly. These are termed as “reference spots”[3]. In the present study we use the breeding site as the main reference spot to further analyze the effect of environmental variables on behaviour of the gulls.

Environmental Data

To test the application of cross wavelet analysis for mining periodic behavior of the gulls in relation to environmental co-variates, we use the movement parameter, step length during the breeding season, as well as three environmental variables: temperature, precipitation and wind speed. Environmental data was downloaded from the Royal Netherlands Meteorological Institute, Station Vlieland. Mean hourly data for temperature, precipitation and wind speed were used for the study.

Cross wavelet analysis

The wavelet transform of a time series decomposes the data using a wavelet function resulting in a time-scale representation of a temporal pattern [4]. Scale in a wavelet analysis is generated by contraction and dilation of the wavelet function. Contraction or dilation changes the time window over which the wavelet function is applied on the time series. Increasing the size of the window increases the scale at which the wavelet coefficients are calculated. The parameter scale in the wavelet analysis can be paralleled to scales used in maps. Corresponding to high scales in maps, in wavelet analysis higher scales (or periods) represent global view of the signal that usually spans the entire signal and low scales (or periods) correspond to a detailed view that relatively lasts a short time. Thus wavelet coefficients are obtained for a series of scales and at each time stamp. We use the Morlet wavelet for this study defined as

$$\psi(\eta) = \pi^{-1/4} \exp(-i\omega_0\eta) \exp(-{\eta^2}/{2}) \qquad \text{Equation 1}$$

The continuous wavelet transform of the discrete time series $X_n$at scale a and time t_i is defined by

$$W\,[a, t_j] = \frac{1}{\sqrt{a}} \sum_{l=1}^{N} X_j \;\; \psi * [\frac{(l-j)\Delta t}{a}] \qquad \text{Equation 2}$$

where ψ* denotes the complex conjugate of the analyzing wavelet function ψ.

A local power spectrum of a wavelet transform is the square of the wavelet coefficient $|W_n^X\,(s)|^2$for each scale and at each time stamp. Cross wavelet transform (XWT) of two time series x_n (movement parameter) and y_n(environmental parameter) can be defined as

$$W^{XY} = W^X\, W^{Y\,*} \qquad \text{Equation 3}$$

where * indicates the complex conjugate. As the cross wavelet transform gives complex values, it can be decomposed into amplitude and phase angles. The phase angles describe the delay between the two signals at a time t on a scale s.

Thus cross wavelet transform exposes regions with high common power and correlation between the two time series and further reveals information about the phase relationship. The circular mean of the phase angles can be used to quantify the phase relationship with in phase shown by arrows pointing right, anti-phase pointing left, x leading y by 90 degree pointing down and y leading x by 90 degree pointing up.

Regions of significant scalogram values were defined by the “area wise test” [5] which is a bootstrapping test. Regions of modulus values greater than or equal to the 0.95 sample quantile of 1000 bootstrapped coefficient values against a red noise null model fit to the data were considered significant and marked in black contour lines [1, 5-8]. Cross wavelet software was provided by Aslak Grinsted (2002-2004) and used in the Matlab environment.

Results

In Figure 1 cross wavelet analysis of step length and temperature, precipitation and wind speed reveals strong diurnal cycles with specific periods of high common power. For example step length and temperature show consistent high powers at a scale or period of 24 hours, 6 and 24 hour cycles with precipitation and 24 hour cycles with wind speed. For two time series to be related, areas with consistent or slowly varying phase angles should be considered in the regions of high common power. The phase angles give an indication of the lag between the two time series. Step length when analyzed with temperature in the 24 hour period, both time series are in-phase over the entire breeding season. This suggests that a daily activity of one cycle per day is a result of daily oscillation in temperature. The 5% significant areas of high power are phase-locked suggesting a strong link between the two time series. For shorter cycles (periods), the phase angles are not consistent and show that that hourly movement is probably not significantly dependent on the temperature and is affected by other factor. The average phase angle for the 24 hour period is 24 ± 15° and proves that step length of movement essentially mirrors temperature and does not have a significant lag between them.

Cross wavelet analysis of step length with precipitation in Figure 1(b) showes an average angle of 169 ±15°demonstrating that movement and precipitation are predominantly in anti-phase. This indicates that movement and precipitation are inversely related to each other. Cross wavelet analysis between step length and wind speed showed that movement was mainly in-phase at a period of 24 hours. However at higher powers (64 hours or 3 days) the phase angles gradually varies from in phase to anti phase.

Figure 2(a) shows individual variations using cross wavelet analysis of step length with temperature. Consistent high powers from the cross wavelet analysis indicates diurnal cycles in all individuals but also shows heterogeneity in time periods of movement bouts and resting phases. The power spectral values from cross wavelet analysis of all individuals were stacked into a 3D matrix and averaged for each time stamp and at all scales in Figure 2(b). The result shows that correlation between distance moved and temperature was not continuous across time for the temporal scale 24, thereby showing a distinct pattern in the sample under study that can be further linked to other ecological and social factors.

Conclusion

The cross wavelet analysis is used for exploratory data analysis as a first step towards modeling moving animals. The traditional methods of visually analyzing trajectories of animals tagged with telemetry devices are replaced with mathematical models and statistics. Advanced data mining techniques can process large amounts of telemetry data in a conceptually standardized way. The method proposed serves as the first step towards understanding the organism-environment interaction, and is the basis for choosing the right environmental variables at the right scale for detailed modeling.While the study uses only a few environmental variables, it illustrates that it can easily be replicated to test for a large range of variables and in doing so we can identify parameters that are capable of explaining cyclic behaviors. This is essential for segregating complex variables into categories such as limiting factors, or into factors that affect the cognitive senses of the animal.

Most studies that examine animal behavior with environmental variables are not able to account for the fact that different components of the environment are important to animals at different scales (temporal scales). Identifying when or for how long a particular environmental cue triggers a specific behavior is the first step towards explaining why and when animals move to certain parts of their range. However this important step is often skipped due to lack of sufficient data at the right scale or tools to identify the scales according to the species under study. Incorrect definition of scales, relative to the perception of space and time by an animal may result in the failure to measure response to variables and variation relevant to the process of interest.

Ethical statement and acknowledment

In order to catch and fix transmitters a license of the “Flora en Fauna Wet”, number FF75A/2007/056 and approval from the Dutch Ethical Committee under protocol number CL 0703 was obtained for the gulls. This project was performed thanks to the bird migration data collected through FlySafe (http://iap.esa.int/flysafe), a project of the European Space Agency Integrated Applications Promotion (IAP) programme (http://iap.esa.int/). This work was supported by the European Union Erasmus Mundus Programme (2008-3620/001-001-MUN ECW) PhD award to Maitreyi Sur.

a) Cross wavelet analysis of step length and temperature

b) Cross wavelet analysis of step length and temperature

c) Cross wavelet analysis of step length and wind speed.

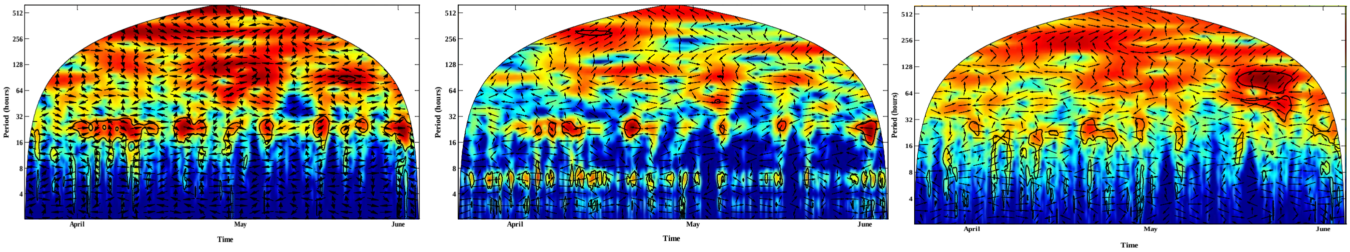


Figure 1. Cross wavelet transform of time series of movement parameter step length and environmental variables. The 5% significant level against red noise is shown as a thick contour. Larger squared modulus values correspond to warmer colours (red, yellow) and smaller values correspond to cooler colours (blues).The relative phase relationship is shown as arrows.

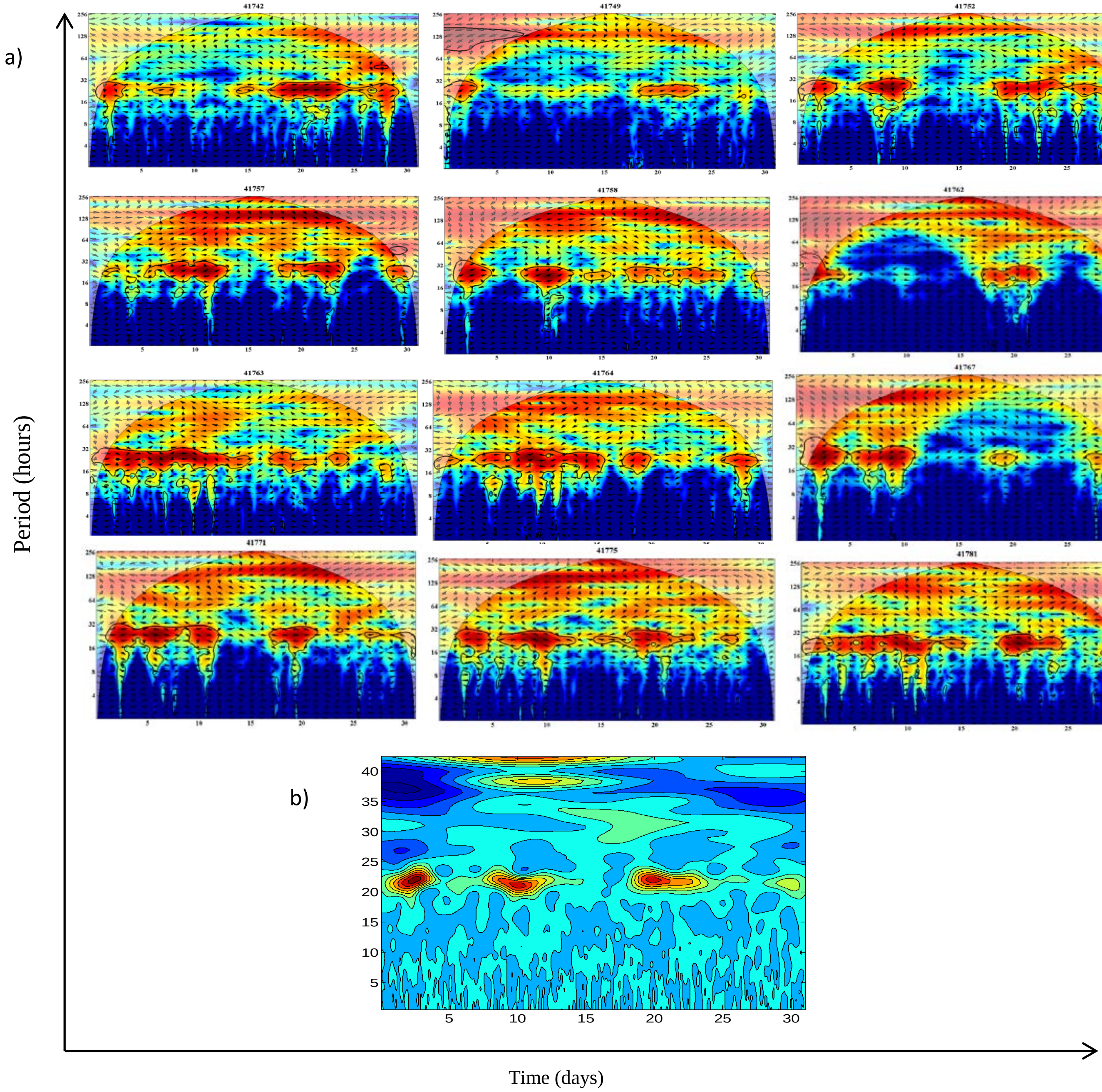


Figure 2. a) Cross wavelet analysis of step length and temperature of twelve gulls during the breeding season (May2008). b) Average power spectral values from cross wavelet analysis of step length and temperature of all the individuals.

References

[1] Polansky, L. et al., (2010). From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology, **91**(5)*, 1506-1518.

[2] Ropert-Coudert, Y., Wilson, R.P. (2005). Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment, **3**(8)*, 437-444.

[3] Patterson, T.A., et al. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology, **78**(6)*, 1113-1123.

[4] Gaucherel, C. (2011). Wavelet analysis to detect regime shifts in animal movement. *Computational Ecology and Software, **1**(2)*, 69-85.

[5] Maraun, D., Kurths, J., Holschneider, M. (2007). Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing. *Physical Review E, **75**(1)*, 016707.

[6] Maraun, D., Kurths, J. (2004). Cross wavelet analysis: significance testing and pitfalls. *Nonlinear Processes in Geophysics, **11**(4)*, 505-514.

[7] Torrence, C.,Compo, G.P. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society, **79**(1)*, 61-78.

[8] Bartumeus, F. (2009). Behavioral intermittence, Lévy patterns, and randomness in animal movement. *Oikos, **118**(4)*, 488-494.

Time-Place Learning in Mice: A circadian system dependent learning and memory task

C.K. Mulder[1,2*], M.P. Gerkema[2], E.A. van der Zee[1]

[1] Department of Molecular Neurobiology, University of Groningen, Groningen, the Netherlands,
[2] Department of Chronobiology, University of Groningen, Groningen, the Netherlands
* cornelis.mulder@rug.nl

Time-Place Learning (TPL) is the process in which animals link biological significant events (e.g. encountering predators, food, mates) with the location and time of occurrence in the environment. This allows animals to anticipate which locations to visit or avoid based on previous experience and knowledge of the current time of day. Recent advances have increased our knowledge on establishing TPL in a laboratory setting, leading to the developed a now well established TPL paradigm in a three-arm maze for mice [1, 2], see Figure 1. This paradigm reflects the natural situation in which foraging animals have to evaluate risks connected with different feeding locations, which may be safe or unsafe to visit depending on the time of day.

Figure 1. The TPL maze. Mice are food deprived to 85% of ad libitum feeding weight, but can find food at the end of each of the maze arms, behind a small metal grid. In each of the 3 daily sessions (lasting maximally 10 min per mouse), mice have to learn to avoid 1 of the 3 feeding locations, depending on the time of day (i.e., session). On visiting a 'wrong' location, mice receive a mild but aversive foot shock (<1 s).

Using this paradigm, we previously demonstrated that, rather than using external cue based strategies, mice use an internal clock for TPL. This finding has made TPL an interesting research topic from several perspectives: 1: TPL is a suitable task to study the role of the circadian system in associative memory formation (with time of day as a discriminative contextual cue). 2: TPL enables the investigation of circadian clock components on a functional behavioral level. First studies have shown that TPL is *Cry* clock gene dependent, but independent of *Per* clock genes [3]. 3: TPL offers the possibility to study the functional interaction between learning/memory and the circadian system with aging. First results have shown that old mice (>17 months) use external cues instead of an internal clock for TPL. We found no significant aging effect in a spatial memory test (spontaneous alternation), but found significant aging effects in circadian system behavioral parameters. These data suggest that at old age, the circadian system is too weak for TPL and mice are forced to use an alternative (suboptimal) strategy [4]. Finally, the pros and cons of this paradigm to measure timing behavior in mice will be discussed, as well the development of an automated TPL set up.

References

[1] Mulder, C.K., Gerkema, M.P., Van der Zee, E.A. (2013). Circadian clocks and memory: time-place learning. *Front Mol. Neurosci* **6**, 8.
[2] Van der Zee, E.A., Havekes, R., Barf, R.P., Hut, R.A., Nijholt, I.M., Jacobs, E.H., Gerkema, M.P. (2008). Circadian time-place learning in mice depends on Cry genes. *Curr. Biol.* **18**, 844-848.
[3] Mulder, C.K., Van der Zee, E.A., Hut, R.A., Gerkema, M.P. (2013). Time-place learning and memory persist in mice lacking functional per1 and per2 clock genes. *J. Biol Rhythms* **28**, 367-379.
[4] Mulder, C.K., Gerkema, M.P., Van der Zee, E.A. (2013). Individual differences in time-place learning: comparison of life-long training with first-time training at old age. *20th international KNDV Zoology conference on individual differences* (Groningen, 1-3 November 2013), poster presentation.

Activity pattern relates to body temperature variation of differently sized South African antelopes

Frank van Langevelde, Anil Shrestha, Sip van Wieren, Steven de Bie, Herbert Prins

Resource Ecology Group, Wageningen University, Wageningen, The Netherlands

frank.vanlangevelde@wur.nl

Heat stress can limit the activity time budget of ungulates due to hyperthermia, which is relevant for African antelopes in ecosystems where temperature routinely increases above 40 °C. Body size influences this thermal sensitivity as large bodied ungulates have a lower surface area to volume ratio than smaller ungulates, and therefore a reduced heat dissipation capacity.

We first tested whether the activity pattern during the day of three antelope species of different body size—eland, blue wildebeest and impala—is negatively correlated with the pattern of black globe temperature (BGT) [1]. Furthermore, we tested whether the larger bodied eland and wildebeest are less active than the smaller impala during the hottest days and seasons. To understand this adaptive behaviour, we investigated daily and seasonal variation in body temperature (Tb) of these three species of antelope, using abdominally-implanted temperature data loggers [2].

The study was conducted at two climatically contrasting environments in South Africa, one with a less seasonal and mild winter (Mapungubwe National Park) and the other with a more seasonal, long and cold winter (Asante Sana Game Reserve) [1,2]. Since the habitat with long and cold winters would be suboptimal for these African antelopes, which evolved in less seasonal and hot environments, antelopes in Asante Sana were expected to exhibit a larger amplitude in Tb and a lower minimum body temperature (MinTb) during winter to reduce Tb and the ambient temperature (Tb-Ta) gradient to save energy.

Our results show that BGT was negatively correlated with the diurnal activity of eland, wildebeest and impala, particularly during summer. During spring, only the activity of the larger bodied eland and wildebeest was negatively influenced by BGT, but not for the smallest of the three species, the impala. We argue that spring, with its high heat stress, coupled with poor forage and water availability, could be critical for survival of these large African antelopes.

In both eland and impala, 24-hour body temperature amplitude did not differ between the study sites, regardless of season. Conversely, wildebeest in Mapungubwe showed a higher variability in the 24-hour amplitude of body temperature and also a lower MinTb during winter and spring than the wildebeest in Asante Sana. This variation in Tb among Mapungubwe wildebeest was influenced by both the amplitude of ambient temperature (positive) and cumulative rainfall (negative), which was not the case for wildebeest in Asante Sana. We propose that the low MinTb of wildebeest in Mapungubwe was the result of nutritional stress during winter and spring; an evident response even during a year of average rainfall. Therefore, these wildebeest apparently live in a physiologically stressful environment.

With the predicted increase in the frequency and intensity of drought periods in southern Africa, wildebeest, and other grazers, will likely experience greater nutritional stress in the future. Our study contributes to understanding how endothermic animals can cope with extreme climatic conditions, which are expected to occur more frequently due to climate change.

All experimental procedures were approved by the Animal Ethics Screening Committee of the University of Witwatersrand (protocol no. 2007/60/4) and South African National Park.

References

[1] Shrestha A.K., Van Wieren, S.E., Van Langevelde, F., Fuller, A., Hetem, R.S., Meyer, L.C.R., De Bie, S., Prins, H.H.T. (2014). Larger antelopes are sensitive to heat stress throughout all seasons but smaller antelope only during summer in an African semi-arid environment. *International Journal of Biometeorology* **58**:41-49

[2] Shrestha A.K., Van Wieren, S.E., Van Langevelde, F., Fuller, A., Hetem, R.S., Meyer, L.C.R., De Bie, S., Prins, H.H.T. (2012). Body temperature variation of South African antelopes in two climatically contrasting environments. *Journal of Thermal Biology* **37**:171-178

Inhibitory Avoidance Learning in Zebrafish (*Danio rerio*)

Remy Manuel[1], Marnix Gorissen[1], Lars Ebbesson[2], Jan Zethof[1], Hans van de Vis[3], Gert Flik[1] and Ruud van den Bos[1]

[1]Department of Organismal Animal Physiology, Institute for Water and Wetland Research, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
[2]Uni Research AS, Thormøhlensgt. 49B, N-5006 Bergen, Norway
[3]IMARES, Wageningen UR, P.O. Box 77, 4401 NT Yerseke, The Netherlands

Introduction

The zebrafish (*Danio rerio*) is increasingly being used as model in behavioural [1,2], neurobiological and genetic research [3]. Underlying reasons are high genetic homology to humans [4] and the many advantages over the use of rodents, such as low cost, easy handling, short reproduction cycle and high fecundity. Furthermore, its genome [5], transcriptome [6] and proteome [7] are well described, making the species a model of choice for behavioural research linked to genetics.

An emerging field addresses learning and memory related to anxiety and fear behaviour [8,9], which has been studied through inhibitory avoidance paradigms [10-12]. Assessment of inhibitory avoidance learning in zebrafish is based on the conflict that arises when a fish wants to enter a dark area to avoid a brightly lit area (innate response; innate anxiety) and avoid this dark area, as it has been associated with an electric shock as negative stimulus. Higher latencies of entering the dark area are indicative of increased inhibitory avoidance learning. Similar paradigms have been studied extensively in rodents. The number of studies on this paradigm in zebrafish is limited.

We therefore studied inhibitory avoidance learning in a series of experiments and assessed whole-body cortisol content and telencephalic expression levels of task-relevant genes. Figure 1 summarises the task and subject variables as well as read-out parameters in relation to the inhibitory avoidance paradigm

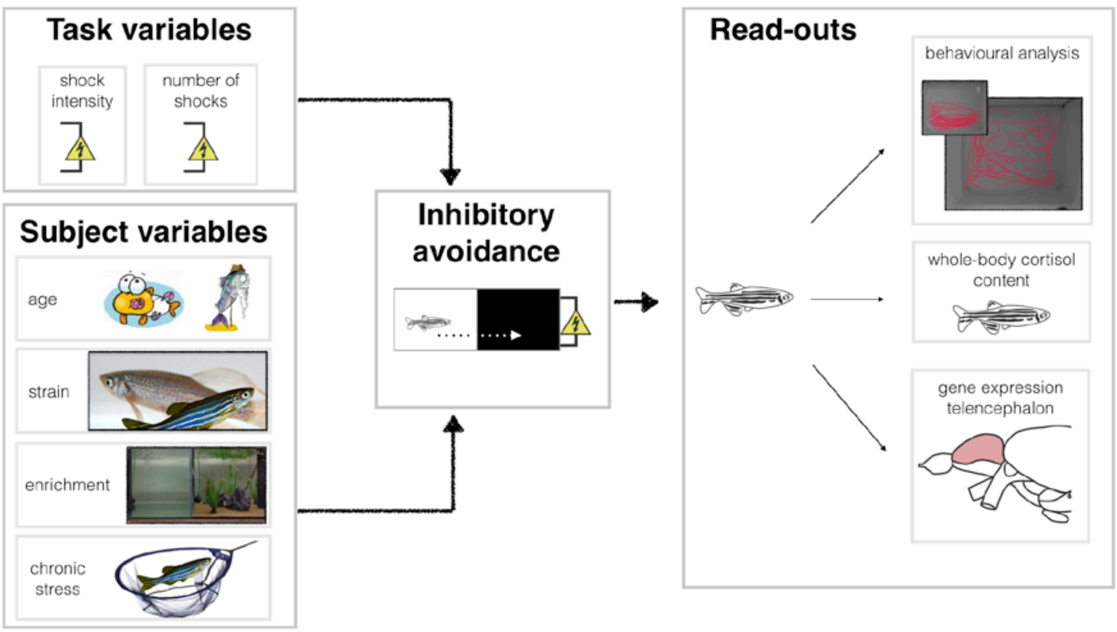


Figure 1. Schematic overview of the task and subject variables as well as read-out parameters in relation to the inhibitory avoidance paradigm.

Material and Methods (general)

Zebrafish and housing

For our experiments we used an in-house-reared Tupfel Longfin (also known as Tübingen longfin; Aquarium facilities of the Radboud University Nijmegen) zebrafish strain. Animals were housed according to standard procedures. Four weeks before inhibitory avoidance learning, fish were housed as experimental groups (approx. 20 fish) in one of two equally sized compartments (50 · 50 · 60 cm) created by separating an aquarium (100 · 50 · 60 cm) in two halves by a blackened glass plate. In each aquarium the water level was brought to 30 cm, to realize a total water volume of 75 L per compartment. The two compartments would share a single biological filter (300 L volume) over which the inflowing water had passed. Each compartment received an aeration stone and was provided with independent water inflow and outflow (26°C, pH 8.0); The photoperiod was kept at 12L:12D (lights on from 07:00 to 19:00 hour) with feeding moments at 09:00 hour (artemia) and 15:00 hour (TetraMin; Tetra, Melle, Germany).

Inhibitory avoidance paradigm

Inhibitory avoidance learning was assessed between 11.00 and 13.00 hours under normal light conditions (400 lux; measured at the water surface). The inhibitory avoidance tank was as follows: an aquarium (60 · 30 · 30 cm; 10 cm water level) was split into two equal compartments, separated by a manually operated sliding door. Surfaces of one compartment were white, while surfaces of the other compartment were black. Compartments were not covered by a lid. The black compartment contained two metal plates (covered with black sound box mesh to prevent light reflection) that covered two opposite walls completely. Both plates were wired to a power source that allowed for an electric current to be put between the plates in the water.

On each day of the inhibitory avoidance paradigm, fish were individually placed in the white compartment with the sliding door closed. After 60 s (acclimation period) the sliding door was lifted, giving the fish free access to the black compartment. Once the fish had completely entered the black compartment, the sliding door was closed and an electric shock was given for 5 s after which the fish was returned to its home tank or no shock was given and fish were sampled, depending on the day of the experiment. Fish were given a total of 180 seconds to enter or avoid the black compartment.

Operation of the sliding door was done by hand; visual observation was done in real-time and recorded on-site. Recording of latency times was done by stopwatch. All procedures were carried out in a manner that caused the least possible disturbance of fish during the experiment.

Behavioural observations were recorded by camera for the white compartment only. Recordings were analysed *post-hoc* using EthoVision XT v10.0 tracking software (Noldus, Wageningen, the Netherlands). We assessed fish behaviour both before and after opening of the sliding door.

Ethical approval

Experimental procedures were approved by the ethical committee of the Radboud University Nijmegen and were conducted in line with Dutch law (Wet op de Dierproeven 1996) and European regulations (Directive 86/609/ EEC).

Results

Our data show that manipulation of task-related features, such as increasing shock intensities [13] or the number of shocks [14], increased inhibitory avoidance learning. Furthermore, the effects of environmental manipulations on inhibitory avoidance learning, such as environmental enrichment [15] and unpredictable chronic stress in combination with changes in photoperiod [14], both showed reduced inhibitory avoidance learning. In addition, inhibitory avoidance learning was strongly reduced in aged (24-month old) zebrafish [15]. Recently, we observed strong strain-related (AB strain *versus* Tupfel longfin strain) differences in inhibitory avoidance learning [16]. In all of these studies we observed two distinct behavioural types. Fish that did not enter the black compartment (latencies > 180 s), marked as avoider fish, and fish that entered the black compartment (latencies < 60 s), marked as nonavoider fish. A selection of these studies will be presented in greater detail, showing how behavioural data may be linked to differences in telencephalic gene expression and whole body cortisol content.

Conclusion

The data suggest that the zebrafish is a suitable model to study processes underlying inhibitory avoidance learning. In addition, these data serve as starting point for unravelling individual differences in brain-behaviour relationships underlying learning and memory related to anxiety and fear in zebrafish.

Reference list

[1] Miklósi, Á., Andrew, R.J. (2006). The zebrafish as a model for behavioral studies. *Zebrafish*, ***3***, 227-334

[2] Gerlai, R. (2010). High-throughput behavioral screens: the first step towards finding genes involved in vertebrate brain function using zebrafish. *Molecules*, ***15***, 2609–22.

[3] Lieschke, G.J., Currie, P.D. (2007). Animal models of human disease: zebrafish swim into view. *Nature Reviews Genetics*, ***8***, 353–67.

[4] Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, ***496***, 498–503.

[5] Jekosch, K.. (2007). The zebrafish genome project: Sequence analysis and annotation. *Zebrafish: 2nd Edition Genetics Genomics and Informatics*, ***77***, 225–39.

[6] Mathavan, S., Lee, S., Mak, A., Miller, L.D., Murthy, K., Govindarajan, K.R., et al. (2005). Transcriptome analysis of zebrafish embryogenesis using microarrays. *Plos Genetics*, ***1***, 260–76.

[7] Lucitt MB, Price TS, Pizarro A, Wu W, Yocum AK, Seiler C, et al. (2008). Analysis of the zebrafish proteome during embryonic development. *Molecular & Cellular Proteomics*, ***7***, 981–94.

[8] Stewart A, Gaikwad S, Kyzar E, Green J, Roth A, Kalueff AV. (2012). Modeling anxiety using adult zebrafish: A conceptual review. *Neuropharmacology*, ***62***, 135–43.

[9] Kalueff AV, Stewart AM, Kyzar EJ, et al. (2012) Time to recognize zebrafish “affective” behavior. Behaviour, ***149***, 1019–36.

[10] Blank M, Guerim LD, Cordeiro RF, Vianna MRM. (2009) A one-trial inhibitory avoidance task to zebrafish: rapid acquisition of an NMDA-dependent long-term memory. *Neurobiology of Learning and Memory*, ***92***, 529–34.

[11] Piato ÂL, Capiotti KM, Tamborski AR, Oses JP, Barcellos LJG, Bogo MR, et al. (2011). Unpredictable chronic stress model in zebrafish (Danio rerio): Behavioral and physiological responses. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, ***35***, 561–7.

[12] Ng M-C, Hsu C-P, Wu Y-J, Wu S-Y, Yang Y-L, Lu K-T. (2012). Effect of MK-801-induced impairment of inhibitory avoidance learning in zebrafish via inactivation of extracellular signal-regulated kinase (ERK) in telencephalon. *Fish Physiology and Biochemistry*, ***38***, 1099–106.

[13] Manuel, R., Gorissen, M., Zethof, J., Piza Roca, C., van de Vis, H., Flik, G., Bos, van den, R. (2014). Inhibitory avoidance learning in zebrafish (Danio rerio): effects of shock intensity and unravelling differences in task-performance. *Zebrafish*. doi: 10.1089/zeb.2013.0970

[14] Manuel, R., Gorissen, M., Zethof, J., Ebbesson, L.O.E., van de Vis, H., Flik, G., & Bos, van den, R. (*Submitted*) Unpredictable chronic stress decreases inhibitory avoidance learning in Tuebingen Long-Fin zebrafish (*Danio rerio* Hamilton): stronger effects in the resting phase than in the active phase. *Journal of Experimental Biology*.

[15] Manuel, R., Gorissen, M., Zethof, J., Ebbesson, L.O.E., Stokkermans, M, van de Vis, H., Flik, G., Bos, van den, R. (in prep.) The effect of enrichment and age on inhibitory avoidance in zebrafish (*Danio rerio)*

[16] Gorissen, M., Manuel, R., Wolf, de, M., Mes, W., Flik, G., Bos, van den, R. (*in prep.*) Differences in inhibitory avoidance learning between two strains of zebrafish (*Danio rerio*): AB *versus* Tupfel longfin

Biofeedback protocols for stress management in high-performance work environments

M. Munafò[1,2], D. Palomba[1], and M. Casarotti[1,2]

[1] Department of General Psychology, University of Padova, IT
[2] INSIDE SRL, Padova, IT

Abstract

High demanding work environments are sources of potentially harmful stress reactions. Chronic stress experienced in working life generates an autonomic dysregulation, which in turn results in high resting heart rate and blood pressure, as well as reduced vagal tone. Here, we proposed a novel approach that exploits competitive biofeedback to reduce negative psychological and physiological outcomes of stress in high-performance work environments. We compared standard to competitive biofeedback, showing that competition enhances training efficacy in promoting health and restoring cardiac autonomic balance.

Keywords: stress management, biofeedback, cardiovascular risk, psychophysiology, respiratory sinus arrhythmia, competition

Introduction

Work environments with high demands are sources of potentially harmful stress reactions. Stress and job strain drive people to adopt unhealthy lifestyles (smoking, low physical activity, high-calorie diet, excessive alcohol consumption) and influence the activity of immune, endocrine, and cardiovascular systems, increasing disease susceptibility and progression across a broad spectrum of disorders [10]. In particular, stress and job strain represent a risk factor for development and exacerbation of cardiovascular diseases [17]. In response to stressors, the autonomic nervous system triggers short-lasting adaptive physiological changes, priming the organism for fighting or fleeing. However, prolonged stress alters autonomic regulation with predominant sympathetic activity and reduced vagal tone. This results in high resting heart rate (HR) and blood pressure (BP), impairing cardiovascular health [9]. Traditional stress management programs use various behavioral techniques, including cognitive procedures, relaxation, meditation, deep-breathing, and exercise protocols [14]. Whereas these techniques aim to relieve psychophysiological symptoms of stress, biofeedback training addresses the source of autonomic balance. Here, we propose a novel approach that exploits biofeedback technology to reduce negative physiological outcomes of stress in high-performance work environments. The approach takes advantage of competition as a motivating factor in challenging individuals to enhance their physiological self-regulation. Competition, indeed, contributes positively to achievement [5], enhancing competitor's intrinsic motivation, improving performance, and fostering the mastery of a skill [4, 13]. Our goal in this study was to assess the effect of competitive biofeedback training on stress managment in comparison to standard (non-competitive) biofeedback training. The ability to cope with stress was indexed by respiratory sinus arrhythmia (RSA), which is a measure of parasympathetic nervous system activity [2] and an index of cardiovascular health [13]. The ability to enhance RSA is associated with physiological flexibility to environmental demands [1] and emotional self-regulation [11].

Method

Participants

Thirty-six managers, recruited from private (banking group, manufacturing industries and media) and public (health service, education system, local government and military) companies with high-competitive work environments took part in the study. Participants were randomly assigned to the Competitive (n = 18) or Standard (n = 18) training group. All partcipants were males, aged 35-67 years, in active employment status, with no history of heart problems or other chronic mental or neurological disease. We did not include women in the study because of the few female managers in the recruiting area. None of the participants were taking medications influencing heart rate (i.e. beta-blockers), tranquilizers or antidepressants. All participants were instructed about the study's procedure and provided their written informed consent. The research was carried out in accordance with the Declaration of Helsinski and the study protocol was approved by the ethic committees of the involved institutions. For data analyses, we included only participants who completed all the training sessions and who had no major training or post-training scheduling irregularities. Five participants did not complete the training schedule, thus the final sample used for the analyses consisted of 31 participants (Age: M = 47.75; SD = 7.87), 17 in the Standard and 14 in the Competitive training group.

Measurements and apparatus

We recorded blood volume pulse (BVP) and abdominal respiration using a FlexComp Infiniti™ encoder and BioGraph Infiniti software (Thought Technology Ltd, Montreal, Quebec). Data were processed via a 14-bit analog-to-digital converter with a sampling rate of 256 Hz (bandwidth DC – 64Hz) and stored for analyses in a IBM-compatible Intel Core™ 2 computer. BVP was measured with a photoplethysmographic sensor (BVP-Flex/Pro) on the right ring finger. Heart rate (HR) was computed as a reciprocal of R-R intervals (RRI) derived from BVP signal. Respiration was measured using a strain gauge sensitive Respiration-Flex/Pro sensor stretched around the participant's abdomen. The software calculated respiration rate from differences in chest expansion in the raw signal waveform. Kubios-HRV 2.0 (Kuopio, Finland) software was used to correct artifacts with a piecewise cubic spline interpolation method. RSA was then computed with a fast Fourier transformation (FFT) on RRI. For each participant, RSA was estimated as the power spectrum of RRI within the range of the individual respiratory rate. HR, abdominal respiration, and RSA were fed back to participants via a 15-inch PC display, positioned in front of them at a distance of 50 cm.

Procedure

All participants received a semi-structured interview, aimed at collecting socio-demographic and health behavior data, and two questionnaires: the Occupational Stress Indicator (OSI) [5], which assesses job stress and its effects, and the Jenkins Activity Survey (JAS) [6], which evaluates Type A behavior. At pre-training and post-training, participants received a psychophysiological assessment. This started with a 4-minute period of rest during which BVP and respiration rate were recorded to derive RSA data. Biofeedback training consisted of 5 weekly sessions, each lasting about 40 minutes. Each session included a 4-minutes resting period, two 6-minutes biofeedback trials, and another 4-minutes resting period. Participants attended training sessions in pairs, matched for age, body mass index, and level of physical activity. In the Standard training group, each participant was trained to synchronize his breathing and HR tachogram (i.e., beats/min) until they covaried in perfect phase, with the goal of maximally increasing the amplitude of RSA. A horizontal bar provided feedback on the individual ability to increase RSA. The bar grew from left to right, by one incremental step, each time RSA exceeded a threshold value for 10 consecutive seconds. The threshold was computed at the beginning of each session, on the basis of the RSA mean value recorded during rest period. In the Competitive training group, each participant was asked to synchronize his breathing and HR tachogram better than the challenger (i.e., the paired participant). The system compared the gain in RSA of both participants each respiratory cycle. Accordingly, a horizontal bar grew, by one incremental step, only for the participant who was more efficient at enhancing RSA for 10 consecutive seconds.

Data Reduction and Analysis

Student's *t* tests for independent groups were performed on mean age, body mass index, sleep time, JAS and OSI subscale scores. χ^2s were calculated for educational qualifications, sleep disorders, smoking habits, physical activity levels, family history of hypertension and cardiovascular diseases. Mean values were computed for RSA and a log transformation was applied for data normalization. A mixed-model analysis of variance (ANOVA) with Group (Competitive or Standard) as the between-subjects factor, and Time (pre-training or post-training) as the within-subjects variable, was performed on mean RSA. Significant main effects and interactions ($p < 0.05$) followed Newman–Keuls' post-hoc comparisons to identify specific differences.

Results

Groups did not differ in any socio-demographic characteristics, health behaviors, JAS or OSI score. Thirty participants (97%) had a Type A percentile score above 55 and twenty-three (74%) presented a score above 75, indicating high impatience, ambition, competitiveness, time urgency, and hostility in the whole sample. A 2 (Group: Competitive, Standard) X 2 (Time: pre-training, post-training) mixed-model ANOVA on mean RSA at rest yielded a significant main effect of Time [$F(1, 29) = 14.67$; $p < .0005$], as well as a significant interaction [$F(1, 29) = 4.44$; $p < .05$]. There was no significant main effect of Group. Post-hoc comparisons revealed a significant increase in RSA from pre to post training (2.01 vs. 2.77 log ms^2) in the Competitive training group only (see Figure 1).

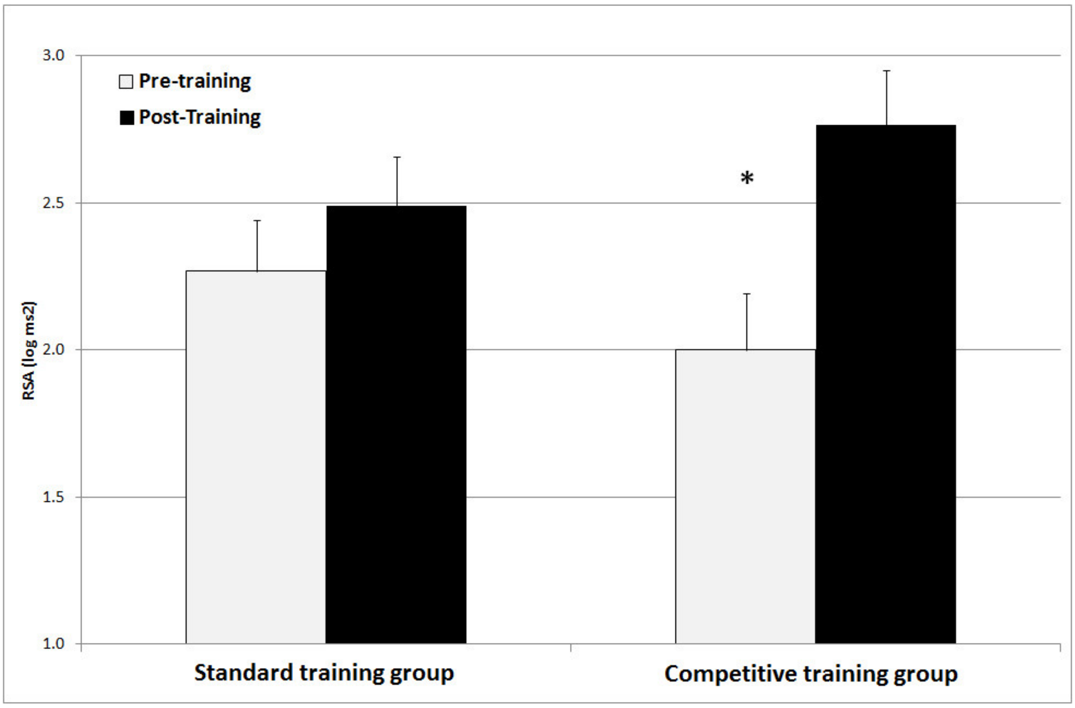


Figure 1. Mean RSA (log ms^2) pre and post-training in the Standard and Competitive training group.

Discussion

The study examined the efficacy of competitive biofeedback training in reducing negative physiological outcomes of stress in high-performance work environments. Our results suggest that competition may be a crucial factor in improving the ability to enhance RSA. Indeed, only the group trained with competitive biofeedback showed a significant RSA increase after training, regardless of the outcome of competition. RSA is an index of parasympathetic predominance over sympathetic activity. Increasing RSA fosters cardiac autonomic balance, that in turn is associated with physiological flexibility and adaptive regulation [1, 8]. Given our small sample size, however, further research is necessary to clarify the effects of competitive biofeedback in the acquisition of autonomic regulation. Moreover, follow-up studies are required to investigate whether the positive effects of competitive biofeedback on vagal cardiac regulation are long-lasting. Competition, indeed, may foster extrinsic motivation, giving negative outcomes in the long run. Our findings are consistent with studies suggesting that, when striving to excel, competitive persons could be motivated for better performance even when this requires reducing physiological activation [12, 16, 18]. Our sample, indeed, presented strong Type A behavior traits, such as impatience, ambition, and competitiveness. Future studies are needed to investigate whether competition in non-Type A individuals facilitates self-regulation in the same way or, on the contrary, prevents a generalized sympathetic withdrawal. In conclusion, the present study suggests that highly competitive individuals might profit from competitive biofeedback training for improving self-regulation strategies, restoring cardiac autonomic balance and reducing the risk for cardiovascular disease.

References

[1] Appelhans, B., Luecken, L.J. (2006). Attentional processes, anxiety, and the regulation of physiological reactivity. *Anxiety, Stress, and Coping* **19**, 81-92.

[2] Berntson, G.G., Cacioppo, J.T., Quigley, K.S. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* **30**, 183–196.

[3] Butler, E.A., Wilhelm, F.H., Gross, J.J. (2006). Respiratory sinus arrhythmia, emotion, and emotion regulation during social interaction. *Psychophysiology* **43**, 612–622.

[4] Cooke, A., Kavussanu, M., McIntyre, D., Ring, C. (2011). Effects of competition on endurance performance and the underlying psychological and physiological mechanisms. *Biological Psychology* **86**, 370–378.

[5] Cooper, C.L., Sloan, S J., Williams, S. (1988). *Occupational Stress Indicator: Management guide*. Windsor: NFER-Nelson.

[6] Jenkins, C. D., Zyzanski, S. J., Rosenman, R. (1979). *Jenkins Activity Survey*. Manual. New York: The Psychological Corporation.

[7] Lam, S.F., Yim, P.S., Law, J.S.F., Cheung, R.W. (2004). The effects of competition on achievement motivation in Chinese classrooms. *British Journal of Educational Psychology* **74**, 281–296.

[8] Lehrer, P.M., Woolfolk, R.L. (2007). Research on clinical issues in stress management. In P.M. Lehrer, R.L. Woolfolk, W.E. Sime (Eds.), *Principles and practice of stress management* (3rd ed., pp. 703–721). New York: Guilford.

[9] Lucini, D., Di Fede, G., Parati, G., Pagani, M. (2005). Impact of chronic psychosocial stress on autonomic cardiovascular regulation in otherwise healthy subjects. *Hypertension* **46**, 1201-1206.

[10] McEwen, B.S. (1998). Protective and damaging effects of stress mediators. *New English Journal of Medicine* **338**, 171–179.

[11] Oveis, C., Gruber, J., Haidt, J., Cohen, A.B., Shiota, M.N., Keltner, D. (2009). Resting respiratory sinus arrhythmia is associated with tonic positive emotionality. *Emotion* **9**, 265–270.

[12] Palomba, D. Stegagno, L. (1993). Competizione, biofeedback e variazioni della frequenza cardiaca. In A.M. Negri Dellantonio E. Mainardi Peron (Eds.), *Psicologia della motivazione: indicazioni di ricerca*. CLEUP, Padova.

[13] Porges, S.W. (2007). The polyvagal perspective. *Biological Psychology* **74**, 116–143.

[14] Richardson, K.M., Rothstein, H.R., (2008). Effect of occupational stress management intervention programs: a meta-analysis. *Journal of Occupational Health Psychology* **13**, 69-93.

[15] Segerstrom, S.C., Solberg Nes, L. (2007). Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological Science* **18**, 275–281.

[16] Shahidi, S., Salmon, P. (1992). Contingent and non-contingent biofeedback training for Type A and B healthy adults: Can type As relax by competing? *Journal of Psychosomatic Research* **36**, 477-483.

[17] Smith, T.W., Ruiz, J.M. (2002). Psychosocial influences on the development and course of coronary heart disease: Current status and implications for research and practice. *Journal of Consulting and Clinical Psychology* **70**, 548–567.

[18] Stern, G.S., Elder, R.D. (1982). The role of challenging incentives in feedback-assisted heart rate reduction for coronary-prone adult males. Biofeedback and Self Regulation 7, 53-59.

[19] Tauer, J.M., Harackiewicz, J.M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. Journal of Personality and Social Psychology 86, 849-861.

[20] Van Zanten, V., De Boer, D., Harrison, L.K. (2002). Competitiveness and hemodynamic reactions to competition. *Psychophysiology* **39**, 759-766.

Capturing mobile experiences: Context- and time-triggered in-situ questionnaires on a smartphone

M. Capatu[1], G. Regal[2], J. Schrammel[2], E. Mattheiss[2], M. Kramer[1], N. Batalas[3] and M. Tscheligi[2,4]

[1] CURE – Center for Usability Research & Engineering, Vienna, Austria
{capatu,Kramer}@cure.at
[2] AIT Austrian Institute of Technology GmbH, Vienna, Austria
{georg.regal,johann.schrammel,elke.mattheiss,manfed.tscheligi}@ait.ac.at
[3] Department of Industrial Design, Eindhoven University of Technology, Eindhoven, Netherlands
n.batalas@tue.nl
[4] ICT&S, University of Salzburg, Salzburg, Austria
manfred.tscheligi@sbg.ac.at

Introduction

Measuring the behaviour and experience of people unobtrusively in different context situations is desirable in order to achieve more valuable results. Questionnaires applied directly in the situation of interest have the advantage that memory effects are prevented, which are a common issue with retrospective measures [1]. However, they also interrupt the ongoing interactive experience. To be able to collect feedback directly in a targeted context situation either requires direct context interviews by a researcher, or the use of mobile devices that the study subject can use in the situation. Context interviews are a reliable in-situ measurement method; but they require a study facilitator to accompany each participant during the trial and therefore are very resource and cost intensive. Mobile devices like smartphones are very useful platforms for in-situ questionnaires, as they allow for an unobtrusive interaction with the user on a day-to-day device [2]. It also allows an easier access to the data for the researchers [2], in contrast to inputting the data from paper questionnaires. However, due to the non-presence of a human observer there is a need to create mechanisms that control the triggering of the questions. Different approaches to solve this problem have been proposed and implemented. Questionnaires can be triggered randomly without knowledge of the context situation (traditional experience sampling), or available data on context (e.g. location, availability of wifi, etc.) can be used (triggered ESM).

In this paper we present a new tool for in-situ questionnaires delivered on a mobile device, which implements the idea of being able to react to the imminent context of people and provides an advanced approach to answering questionnaires in-situ. We also describe the insights we gathered from a qualitative user study with 18 participants and point out the challenges we faced.

Related Work

Intille et al [4] described and developed a tool for context aware experience sampling. The tool uses context-aware sensors to collect feedback from the user only in certain situations, and according to them triggering the questions based on context-awareness helps minimize the interruption annoyance of the ESM. This general idea was further developed by different researchers. Micallef et al. [3] aimed for identifying the most interesting temporal slot to ask questions in a 'ground truth' data collection. To find the optimal moment, they detect if a user was either doing a physical activity (by using accelerometers) or was at a certain location (by using Wi-Fi access points). "MyExperience" [5] is an open source system that captures in situ data on mobile phones and portable devices. It captures device usage data, such as number of phone calls, SMS, used applications and media captured, and phone data, like GPS and Bluetooth and calendar information. This information is used for user experience sampling. Lathia et al. [7] developed an open source smartphone library for computational social science. Similar to our approach they used raw sensor data for triggering experience sample questions. Recent research has also found that the selection of trigger sensors and stimuli has an important influence on the data, and that more elaborated methods for defining trigger situations e.g. by combining multiple sensors are desirable in order to increase the quality of the measurement [8].

Design of the In-Situ Tool

In our work we are addressing this problem by developing a modular framework that allows combining and analysing different sources of context information. In our framework we consider different sources of context information: (a) Sensor data that relies directly on the context - lighting conditions, noise conditions, humidity, temperature, etc.; (b) Sensor data that relies on the persons behaviour in this context - heart rate, speed of movement, eye-tracking data etc.; (c) Location information - GPS data, available Wi-Fi access points, cellular signal strength, etc.; (d) device usage data in the context – number of Bluetooth devices arround, number of reveived calls, number of received messages etc. Besides multiple sources of data there is also a wide range of possibilities to react to the context depending on the goal of the study. Typical reactions in research activities are prompting a questionnaire to the user, triggering experience samples, storing the device state, etc. To enable this wide range of possibilities we designed a middleware that receives messages from the context measurement sources, reacts to this message by computing rules and trigger appropriate reaction. This approach is supported by modern software architecture principles. For example the Android operating system relies on (so called) Intents [1] [10] that are used for communication between different user interface screens, but also different applications. Using this message based approach it is possible to combine different modules (e.g. off the shelf applications) to a more complex system. As a proof of concept we developed a context aware experience sampling prototype that follows this modular approach. Basically the system consist of three modules: (1) a framework for measuring context, to be more precise the sensor values of a smartphone (2) a rule engine that reacts to the sensor values and subsequently triggers the desired question (3) a questionnaire app to provide different questions to the user. An overview of the prototypes architecture can be found in Figure 1.

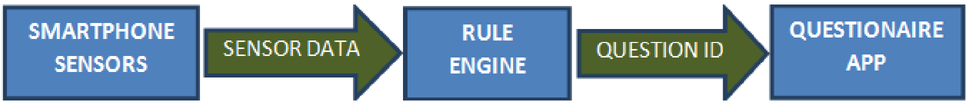


Figure 1. Overview of the approach and the used applications.

Smartphone - Sensors. To measure context information, in our proof-of-concept application we use the built in sensors of an off the shelf smartphone (Samsung Galaxy S4) and additionally an off the shelf heart rate monitor (Zephyr HRM BT). The sensor values were captured using the application AIRS [6]. This app measures selected sensors and subsequently sends intents to our middleware as soon as the sensor value changes. To allow realistic measurement of context parameters (e.g. light or sound) the smartphone was strapped to the user's non-dominant hand – see Figure 2. Also, this placement supports the utilization of the smartphone as input device.

Rule Engine. We developed a middleware that allows the researcher to define rules and thresholds for the sensor measurements which need to be reached, in order to show the user the appropriate questions. The middleware can deal with different rules. Basically there are six types of rules:(1) higher (x > threshold), (2) lower (x < threshold), equals (x=threshold), (4) is not (x != threshold). (5) increase (Δx> threshold) (6) decrease (Δx> threshold). Basic rules can be combined into more complex rules (e.g. if light > 10lumen and noise <100 decibel). Moreover time related thresholds are possible as well (e.g. if noise is higher than 90db for 3 minutes). Additionally it is also possible to set time related mean thresholds (e.g. on average noise is higher than 90db for 3minutes). For our proof of concept implementation a simple approach was chosen. The researcher defines the rules by hand in an XML file which is stored on the device – see Figure 3. Using this approach it is possible to build complex rules that allow detection of complex semantic contexts.

Questionnaire App: The questionnaire app is based on an in-situ data collection tool called Tempest [9] and provides the question to the user after receiving an intent with the question ID from the middleware. It provides different types of questions to the user, e.g. multiple choice questions, free text input, selection grid, etc.

[1] Intent - http://developer.android.com/reference/android/content/Intent.html - Accessed 8. July 2014.

Additionally questions can be combined into a series, which can be triggered by a single intent. All questions are shown and answered on the smartphone.

Figure 2. Smartphone attached to the participants non dominant hand.

```
<CombinedRule>//Type of Rule either CombinedRule or CombinedTimerRule
   <Runs>5</Runs>//How often this rule can be fired
   <QuestionID>QLi1</QuestionID>//Question ID triggered in the questionnaire app
   <BasicRule>
         <ID>ID_MORE</ID>//defines the type of rule - higher, lower, equals, etc.
         <Sensor>com.airs.sensor.LI</Sensor>//the sensor – in this case light
         <Threshold>100</Threshold>// if values is more than threshold
   </BasicRule> // end of first basic rule
</CombinedRule>
```

Figure 3. XML Example - CombinedRule, that consists of one BasicRule. The rule fires if the light sensor is above 100 lumen.

User Study

The goal of the study was to evaluate the practicability of our approach in real conditions and to study the appropriateness of the context detection. Moreover to receive feedback on the perception and opinions of users regarding the context aware questionaire, we compared this approach with time triggered questionaires and retrospective questions. We defined three different context situations a person encounters during a shopping trip in a shopping mall, that are suitable points for measuring user experience: (1) Outside of the building (to measure the experience before/after the overall shopping trip) (2) inside the building (to measure the experience before/after shopping), (3) inside a shop (to measure the experience while actually shopping).

18 participants (nine women) were asked to perform a series of tasks inside two different shops in a shopping mall. The tasks were put in an order that participants passed all three defined situations (outdoor, inside the building, inside a shop). We used a between-subject design with three groups (each consisting of 6 participants). A question triggering approach based on pre-defined time intervals was used for the first group. The second group received questions triggered by the context-aware system. A third group was asked to fill in the questions at the end of the trial, in a retrospective manner (in contrast to the two other groups which answered the questions in-situ as soon as they were triggered). All groups answered the questions on the smartphone. The questions were related to the tasks and featured closed and open-ended questions. After the trial the participants were asked to describe their experience with the used method in a concluding interview.

For the context recognition – based on two pre-studies – we defined the following rules: (1) In shop: Movement speed lower 0.3 km/h, temperature higher 15 degree Celsius and Noise below 70 decibel (2) Indoor: Movement speed between 0.4 and 0.9 km/h, temperature higher 15 degree Celsius and Noise above 65 decibel (3) Outdoors: Movement speed higher 0.8 km/h, temperature lower than 10 degree Celsius (study was conducted in winter).

Results and Discussion

Performance measurements: Within the six participants using the context-recognition setup, there were a total of 42 questionnaire items that were triggered. 37 were triggered correctly, 2 items were triggered in the false context and 3 items were not triggered at all - because the context was not recognized correctly by our system. While the different items were supposed to appear only once during the specific context, and also never twice in a row, there were certain cases were these requirements were not met. E.g. a fast walking speed but long task completion time resulted in the appearance of three questions items in one shop. There was also a problem with the temperature sensor, as it takes a long time to adapt to fast changing temperatures. In the time-triggered group of 42 total questionnaire items triggered, 23 were triggered correctly, 4 only narrowly false and 15 false. The first outdoor questionnaire items were always asked in the correct context. In contrast the first indoor questions afterwards always triggered too early. The question items for inside a shop were triggered false only two times.

User Experience: Participants of all three groups were asked about their experience with filling out questionnaires on a smartphone. The majority of the 18 participants were pleased with answering the questions on the smartphone, though one third said that different input method (especially bigger in size) would have been more helpful. The two groups answering questions in-situ (time triggered and context triggered) said that they felt not distracted by the questions. Participants preferred multiple-choice questions, due to their easiness, but also answering speed. About one third of the participants said that the right answering format is dependent on the situation, as sometimes more detailed and expressive answers are necessary.

13 out of 18 participants felt uncomfortable about carrying the smartphone on their wrist – mainly due to being afraid that it might fall down. More than half of the participants preferred the smartphone to answers on paper. Participants argued that using the smartphone is quicker and more convenient than paper questionnaires. When asked if the questions came at the relevant moment, participants in the context-recognition group said the questions had perfect timing, with the exception of two participants. In the time triggered group, participants had mixed feelings regarding the timing. Some participants perceived that questions were triggered at the right time, even though they arrived too late. The retrospective group was asked, if the questions should have been asked during or right after the task. Participants stated that the short amount of time didn't result in recall difficulties of their opinions and thoughts on the tasks. Although participants hypothesised that if the study duration would have been longer it would have been more accurate in the situation.

Conclusion

In this paper we developed a framework for context aware questionnaires, developed a proof of concept and conducted a first user study. Our work shows that even with a simple rule engine it is possible to react to the user's context accurately. A limitation of our approach is that the rules need to be defined and adapted by hand, which can be tedious. In future work we plan to extend our framework by developing an advanced, intelligent middleware that uses machine learning principles to detect and define appropriate contexts automatically.

References

[1] Hektner, J.M., Schmidt, J.A., Csikszentmihalyi, M. (2007). Experience Sampling Method: Measuring the Quality of Everyday Life. *SAGE Publications, Inc*

[2] Raento M., Oulasvirta A., Eagle N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods and Research, **37**(3),* 426-454.

[3] Micallef, N., Kingdom, U., Just, M., Baillie, L., Kayacik, G. (2013). "Stop Questioning Me !": Towards Optimizing User Involvement during data collection on mobile devices. *Proc. MobileHCI'13, 588–593.*

[4] Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., Bao, L. (2003). A context-aware experience sampling tool. *– CHI EA'03*, 972-973

[5] Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., Landay, J. A., Street, S. B., Mateo, S. (2007). MyExperience : A System for In situ Tracing and Capturing of User Feedback on Mobile Phones, *Proc. MobiSys '07, 57–70.*

[6] Trossen D., Pavel D., (2012). AIRS: A Mobile Sensing Platform for Lifestyle Management Research and Applications. *Proc. Mobilware'12, 1-15*

[7] Lathia, N., Rachuri, K., Mascolo, C., & Roussos, G. (2013). Open source smartphone libraries for computational social science. *UbiComp '13 Adjunct, 911-920*

[8] Lathia, N., Rachuri, K., Mascolo, C., &, Rentfrow, P. (2013). Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods. *Proc. UbiComp '13, 183-192*

[9] Batalas N. & Markopoulos P. (2012). Introducing tempest, a modular platform for in situ data collection. *In Proc. NordiCHI '12, 781-782*

Real-time behavioural analysis using Google Glass

Tengqi Ye, Brian Moynagh, Rami Albatal and Cathal Gurrin

Insight Centre for Data Analytics, Dublin City University, Ireland

Abstract

Lifelogging is a form of pervasive computing that represents a phenomenon whereby people can digitally record their own daily lives in varying amounts of detail, for a variety of purposes. Lifelogging offers huge potential for supporting behaviour change because it can capture the totality of life experience and provide heretofore unknown levels of insight into the real-world activities of the lifelogger. In this paper we present a real-time curated lifelogging prototype that can support real-time behavioural analysis by supporting immediate feedback and intervention to the lifelogger.

Introduction

Lifelogging is "a form of pervasive computing, which generates a unified digital record of the totality of an individual's experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive" [1]. Since lifelogging captures a detailed log of life activity [2], it makes an ideal source of data for behavioural analysis. Indeed lifelogging is already employed for various aspects of behavioural analysis, such as sedentary behaviour [3,4], physical activity recognition [5] and diet monitoring [6,7,8]. This is because a lifelog represents the ultimate "black box" of a human's life activities and as such, the potential mining or inferring knowledge about how we live our lives is enormous [2].

In this paper, we present a real-time lifelogging prototype that operates on Google Glass, in conjunction with server-based architecture that mines/monitors the incoming data to organize, structure and present this data. This lifelogging prototype can operate in both automatic (like a Microsoft SenseCam [9]) or curated modes. In automatic mode, Google Glass will capture and upload multi-sensory lifelog data (including images) at fixed periods without user intervention. In curated mode, the user will choose when to capture lifelog data by means of pre-defined gestures such as an eye-blink. The contribution of this work is that we present the first real-time lifelogging prototype for wearable computing devices (Google Glass) that, with the correct event detectors, could be used to support real-time behaviour analysis and intervention, where real-time typically required a number of seconds between capture and feedback.

Lifelogging for behaviour change

The relationships between lifestyle behaviours and health outcomes are usually based on self-reported data, which is prone to measurement error. Lifelogging has societal applications in terms of providing better fidelity when measuring the behaviour of groups of individuals in a given population, which helps inform policies for tasks like transport planning, environment understanding, and relationships between lifestyle exposures and disease outcomes. Lifelogging sensors, such as wearable cameras and their associated software tools have developed to the point that they are well-suited to measure physical activity, sedentary behaviour, active travel, and nutrition-related behaviours across populations of users [10]. This work fits into this progression by presenting the next-generation real-time lifelogging tool.

From [10], we know that initial lifelogging for behaviour change studies was achieved using devices such as the Microsoft SenseCam [9], which incorporated a VGA camera with fisheye lens, an accelerometer, a light intensity meter, a thermometer and a passive infra-red (PIR) sensor to detect the presence of people. The SenseCam was worn around the neck using a lanyard and by default, a SenseCam captured a new image about every 40 seconds unless triggered by its sensors to capture an image sooner. Captured data was stored using onboard memory, with capacity for about ten days worth of lifelog data. A full day of wearing of a SenseCam would generate between 3,500 and 4,500 images and could support post-capture analysis (i.e. not in real-time). Aside from the SenseCam, there are other dedicated lifelogging wearable devices on the market also, such as the Narrative Clip and the OMG Autographer; both of these operate in a similar manner to a SenseCam. The ubiquity of smartphones suggests that they could also become a valuable real-time lifelogging device and it was recently shown that a smartphone worn on a lanyard around the neck can provide similar levels of lifelogging effectiveness as a dedicated device such as the SenseCam [11], though also support real-time interventions.

Some specific examples of wearable camera lifelogging tools used by public health researchers and others to help inform policy decisions include the work of Kelly et al. [3] who used wearable cameras to identify self-report error in travel behaviour in both adults and adolescents. Kerr et al. [4] employed an annotation framework to manually categorise fine-detail sedentary behaviours from lifelog data, in order to better identify factors that may be driving such behaviour. Doherty et al. [5] were able to identify sedentary, light, moderate, and vigorous intensity physical activities through a combination of accelerometers and wearable cameras. Reddy et al. [6] was able to identify self-reporting errors in a behaviour study of nutrition using a smartphone running lifelogging software. O'Loughlin et al. [7] and Gemming et al. [8] found that they were able to help participants to identify forgotten calories through using wearable camera images as memory prompts. However, in these cases there were two important drawbacks. Firstly, the device did not act in real-time, hence it always provided a retrospective review and could never support real-time interventions. Secondly, wearing a phone or a dedicated device such as a SenseCam on a lanyard means that there is significant potential to miss capturing events of interest, which do not by necessity occur directly in front of the wearer. Hence, having a real-time, head-mounted lifelogging platform offers advantages over any of the previously used devices. Integrating SenseCam type functionality (automatic capture of periodic images, coupled with sensor data) into a Google Glass type device results in a new type of lifelogging device, that tracks head movement, but also supports both automatic or triggered capture, along with the potential for real-time interventions. However, in the current pre-release prototype of Google Glass, battery-power restricts the usage to periods of a few hours at a time. In Figure 1, we show the types of photos that can be captured using a Google Glass device.

Figure 1. The image outputs of a real-time lifelogging tool running on Google Glass, showing social activities, driving and food-related activities.

A real-time lifelogging platform

The real-time lifelogging prototype presented in this paper is based on two core components, a wearable, head-mounted lifelog capture device (software running on Google Glass) and a server-side application that stores, organises, analyses and presents the lifelog data for multi-modal access (See both sides of Figure 2). The data captured by the head mounted device can originate from multiple sensors, including visuals, audio, accelerometer, etc. The list is very similar to a smartphone sensor set; as such the findings in [11] can equally apply to the head-mounted device. Data that is captured on the wearable device gets uploaded to the server immediately, or at the next available opportunity (network availability depending), whereupon it undergoes semantic analysis and organization. The data gathering can operate in either an automatic or curated manner. If it is automatic, then the user does not need to do anything to gather content; the device simply logs photos and sensor readings continually (one every minute, though it is configurable) until the user either switches it off, or until the power is drained[1]. The alternative (or parallel) method of capture is the curated data gathering option which uploads every image explicitly taken (along with appropriate sensor data) to the server for analysis. In curated-mode, photos can be taken in any manner that the device allows; in our case, by button press, in-built voice command or blink-to-capture.

On the server, an initial suite of semantic analysis tools are being developed at present that will operate over the data; these include face detectors, eating detectors, sedentary activity detectors and event segmentation tools. Since these detectors run on the server, there are few CPU and power constraints that would limit the types of analytics that could be performed. An overview architecture is shown in Figure 2, which contains two access devices, three semantic detectors (i.e., faces, eating, sedentary activities) and data stores, data, and interface handlers.

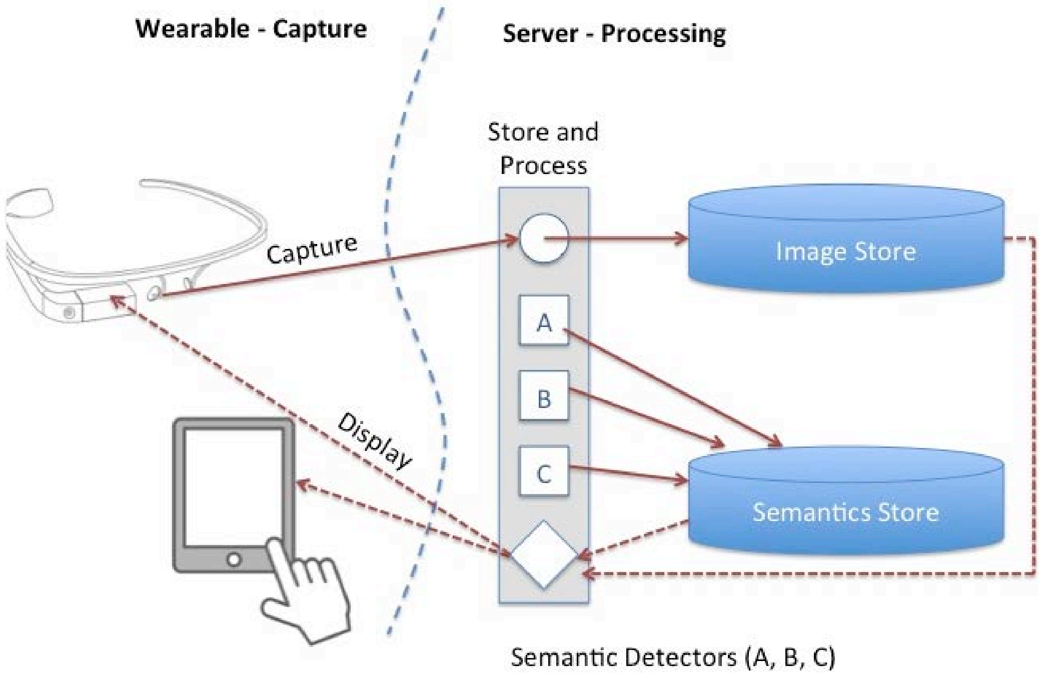


Figure 2. A summary overview of the real-time wearable lifelogging solution.

The access to the processed data (semantic data) could be done using many different types of access device. Access could be via a computer or tablet to support reminiscence or retrospection, or it could be on the capture device itself to support real-time interventions and communication. Naturally the interface elements are heavily dependent on the use cases and the modality of the access devices. For real-time interventions using a wearable

[1] Since Google Glass is currently a pre-release prototype, the battery was not designed to support all-day capture, hence the device can only lifelog continually for a matter of hours.

device such as Google Glass, then the interface should be card based and focus on just the required information at any point in time (e.g. an intervention). For reminiscence or quantified-self style analytics over past life experiences, then a large-screen device that charts, summarise and presents analytics results over the data would be more suitable. Our use-cases below are primarily focused on the reminiscence use cases.

Current & future work

Given a real-time lifelogging platform that can support retrospective access for reminisce or reflection, as well as real-time interventions, there is enormous potential to deploy and evaluate new types of behaviour chance analytics. Some of the use-cases we see big potential for at present include:

- *Digital Diaries* for personal reminiscence and review. Sellen and Whittaker [12] present a suggestion of the five reasons (5Rs) why people would access lifelog stores. One such reason is to support Reminiscence, which would be a key driver of long-term behaviour change and requires the integration of an effective event segmentation technique to organise the lifelog data into a series of events that take place each day.
- *Diet Monitoring*. Gesture-based capture of food being consumed will allow for diet monitoring applications to be developed that can feedback appropriate messages to the user.
- *Product Knowledge*. Object and logo matching in real-time from the point of view of the buyer would allow for immediate analysis of, and feedback based on, the content being purchased.
- *Lifelong Analytics*. Referencing past life experiences potentially over decades opens up new opportunities for behaviour change analysis and intervention.

Of course, these are only a short list of potential use-cases. In reality, there would be a huge list of potential use cases based on analytics tools that understand the user, their activities and environment. In this early work, we focus on the key beneficial use-case of digital diaries, which provide a source of data for either human or automated analytics to support behaviour change. These digital diaries can be explored by the wearer or a third-party, or could even act as input for a suite of analytics tools as described above.

The lifelogging prototype that we present in this work can support real-time analysis and feedback. It is our conjecture that a wearable, head-mounted lifelogging device meets all the same criteria as presented in [11], but does so more effectively by tracking the user head movements, though currently for limited periods of time due to battery constraints. Consequently, we propose that this prototype could be used to gather detailed lifestyle activity records and therefore has potential to be employed as a means of supporting behaviour change.

Acknowledgement

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289

References

[1] Dodge, M., Kitchin, R. (2007). Outlines of a world coming into existence: Pervasive computing and the ethics of forgetting. *Environment and Planning, **34**(3)*, 431–445.

[2] Gurrin, C., Smeaton, A.F., Doherty, A.R. (2014). LifeLogging: Personal big data. *Foundations and Trends in Information Retrieval, **8**(1)*, 1-125.

[3] Kelly, P., Doherty, A.R., Berry, E., Hodges, S., Batterham, A.M., Foster, C. (2011). Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a Pilot Study. *International Journal Behavioural Nutrition Physical Activity*, ***8***, 44.

[4] Kerr, J., Marshall, S. J., Godbole, S., Chen, J., Legge, A., Doherty, A.R., Kelly, P., Oliver, M., Badland, H., Foster, C. (2013). Using the SenseCam to improve classifications of sedentary behaviour in free-living settings. *American Journal of Preventive Medicine, **44**(3)*, 290–296.

[5] Doherty, A.R., Kelly, P., Kerr, J., Marshall, S., Oliver, M., Badland, H., Hamilton, A., Foster, C. (2013). Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *International Journal of Behavioral Nutrition and Physical Activity, **10***, 22.

[6] Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., Hansen, M. (2007a). Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In EmNets'07: Proceedings of the 4th workshop on Embedded networked sensors, p. 13-17, Cork, Ireland. ACM Press.

[7] O'Loughlin, G., Cullen, S., McGoldrick, A., O'Connor, S., Blain, R., O'Malley, S., Warrington, G.D. (2013). Using a Wearable Camera to Increase the Accuracy of Dietary Analysis. *American Journal of Preventive Medicine, **44**(3)*, 297–301.

[8] Gemming, L., Doherty, A.R., Kelly, P., Utter, J., Ni Mhurchu, C. (2013). Feasibility of a SenseCam-assisted 24-h recall to reduce under-reporting of energy intake. *European Journal of Clinical Nutrition, **67**(10)*, 1095–1099.

[9] Hodges, S., Berry, E., Wood, K. (2011). SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory, **7**(19)*, 685–696.

[10] Doherty, A.R., Hodges, S.E., King, A.C., Smeaton, A.F., Berry, E., Moulin, C.J., Lindley, S., Kelly, P., Foster, C. (2013). Wearable cameras in health: The state of the art and future possibilities. *American Journal of Preventive Medicine, **44**(3)*, 320–323.

[11] Gurrin, C., Qiu, Z., Hughes, M., Caprani, N., Doherty, A.R., Hodges, S.E., Smeaton, A.F. (2013). The SmartPhone as a Platform for Wearable Cameras in Preventative Medicine. *American Journal of Preventive Medicine, **44**(3)*, 308–313.

[12] Sellen, A.J., Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM, **53**(5)*, 70–77.

A comparison of human and machine learning-based accuracy for valence classification of subjects in video fragments

Y.H. Holkamp[1], J.G.M. Schavemaker[2]

[1] Faculty of EEMCS, Delft University of Technology, Delft, The Netherlands.
[2] TNO, Delft, The Netherlands.
John.Schavemaker@tno.nl

Introduction

Facial expressions are the primary way to show one's emotional state. Automatic recognition of these cues from video using software allows for various improvements in human-computer interaction, ranging from improved feedback for recommender systems to automatic labeling of movies according to the emotions they induce. A number of affective display databases have been created to aid development in this field. These datasets are frequently available for academic use [1, 2, 3], use picture or video stimuli and range from highly controlled [1, 2] to more natural settings [3]. We observe that methods using these datasets report accuracy figures that leave room for improvement [5].

The contribution of our research is a comparison of the accuracy of valence estimation using facial expressions between human annotators and machine-learning classification when using an open affective computing dataset that uses video to induce emotion. This comparison helps to determine to what extent existing methods can be improved or whether optimal accuracy has been reached for facial expression-based methods. To the best of our knowledge, few comparisons have been made on datasets where video stimuli were used.

In order to make this comparison, we have reproduced the method described by Koelstra and Patras [5] for the case of using facial expressions to estimate the valence experienced by the subject. In contrast to the original approach, we used the Noldus FaceReader software [8] to detect and quantify facial activity using FACS [6] Action Units (AU). Additionally, we have conducted an experiment with 15 users and found that human observers exhibit high inter-rater agreement. Yet, we found that certain video fragments obtained from [1] are difficult to classify, even for human observers. The structure of this paper is as follows; first we describe the methods used. Finally we present our findings and discuss these results and conclude this work.

Method

In our experiments we used the MAHNOB-HCI data set [1], which was created for affect recognition and implicit tagging applications and contains face video recordings, EEG data and more. From the affect recognition part of this dataset, we selected the 24 subjects for whom the full set of data was collected without error. We cut the recordings to match the stimulation part of the experiments. To allow for a comparison between our work and [5], we mapped the subjects' self-assessment ratings to two categories; negative valence (rating 1-5) and positive (rating 6-9). Using this data, we conducted two experiments.

Annotation experiment. In the first experiment, we selected two fragments for random stimuli for each subject in the dataset, resulting in a total of 48 videos. We developed a web-based video annotation system where users were asked to view these videos after providing us with their age, gender and nationality. We presented the users with an explanation of their task, after which each user watched the same four videos. After these items were rated, the remaining videos were presented in random order. The former allowed us to use these sessions as 'gold standard' data to possibly explain outliers whereas the latter ensures that the order of the sessions does not influence the ratings.

Once the user watched a video, she was asked to estimate the emotion of the person they saw in the video using a 5-point Likert-scale, ranging 'negative' to 'positive'. We explicitly instructed the subjects to choose 'neutral' if they could not distinguish an emotion. Furthermore the user was asked to rate how visible the emotion of the subject was in their facial expression using a 5-point Likert scale, ranging from ‘not visible’ to ‘very explicit’.

For our experiments we gathered a group of 4 female and 11 male participants in the age groups 20-29 (n=13) and 30-39 (n=2). They have a Swedish (n=1), French (n=1) or Dutch (n=13) nationality and were asked to anonymously complete our annotation task from their home or office computer as courtesy or for research purposes, no actual rewards were provided. To speed up the annotation process, videos were played at four times the normal speed, which was reported in [10, 11] to have limited impact on detection accuracy. This resulted in an average length of 19.4 seconds (σ=4.9). To keep our subjects focused, we presented the raters with a page that encouraged taking a break at three intervals during the experiment but this was not enforced.

To allow for comparison between the valence estimation performance of humans and software-based estimators, we reduced the responses of our raters to two categories, positive and negative valence, where neutral ratings were included in the negative valence category, as was done in [5]. We then compared the ratings provided by our subjects against the self-assessment ratings provided in the MAHNOB-HCI datasets. In addition to the average human accuracy we also take the majority vote into consideration.

Machine-learning classification experiment. In addition to human estimators, we partially reproduced the method described in [5], which uses facial expressions to estimate the valence. In contrast to their system, we used a support vector machine rather than a Gaussian Naive Bayes classifier and we used the off-the-shelf Noldus FaceReader software to provide us with information regarding Action Unit (AU) activation levels.

The method works as follows; the video fragments were automatically annotated using the FaceReader software, resulting in activation levels at different time intervals. For each entry in the list of EMFACS AU combinations associated with emotion [5], we determined the number of onsets, offsets and the difference between onset and offset strength. Using this setup, we detected 29 combinations, resulting in a total of 29 x 3 = 87 features.

As in [5], we applied recursive feature elimination (RFE) as implemented in [9] to select the features used for classification by training a linear SVM and removing the 10% lowest-weighted features until we reach a predefined number. This number was found using a 10-fold cross-validation on our training set of 6 subjects. In order to determine the classification performance, we then performed a leave-one-fragment-out cross-validation on a per-subject basis, resulting in a SVM trained using 19 fragments from one subject and tested using the 20th.

Results

Overall we obtained the statistics shown in Table 1. Our implementation of the method by Koelstra and Patras obtained an accuracy within 4% of the figures reported in the original work. We observe that the human annotators obtained an accuracy of 71.6%, higher than the 66.2% for the best performing machine-learning approach.

Method	Accuracy
Koelstra & Patras [5][1]	64.0%
Our implementation[1]	66.2%
Majority class[1]	62.6%
Human, average[2]	71.6%
Human, majority vote[2]	75.0%
Majority class[2]	66.6%

Table 1. Valence classification accuracy for two different methods. Majority class performance is included to provide a baseline. Differences are not significant according to pairwise t-tests. [1]Using full dataset (N=480). [2]Using subset (N=48).

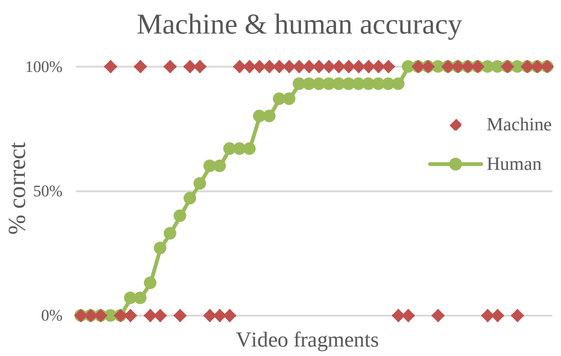


Figure 1. Machine versus human annotator accuracy, the video fragments are presented in ascending order of human accuracy. Note that the machine classifier only rates once and can thus only obtain an accuracy of 0 or 100%.

The differences in ratings between human annotators and the machine learning approach can be seen in Figure 1. We can see that, on the left side of the graph, there is a series of video fragments for which both human and computer fail to provide accurate classification. Conversely, there are nine additional video fragments for which the machine approach fails to provide accurate ratings.

We applied Fleiss' kappa to determine the inter-rater reliability of our human subjects, resulting in an agreement degree of 0.66. According to the index presented by Landis and Koch [7], this corresponds to 'substantial agreement'. We found no significant differences in performance between the different age groups, genders or nationalities among raters.

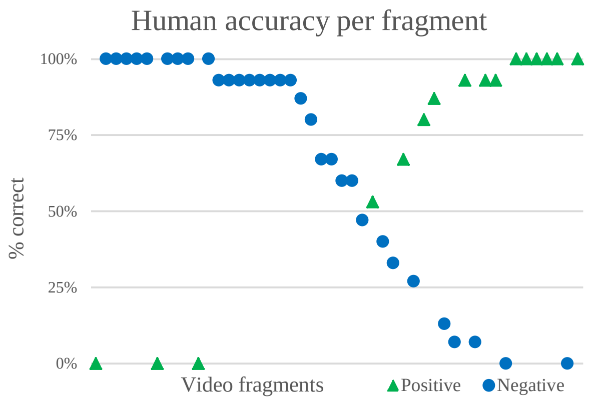


Figure 2. Percentage of correct votes per video fragment. The video fragments were ordered ascending by the number of positive valence ratings. The color and shape show the self-assessment rating provided by the subject in the video.

In Figure 2, we can see the accuracy of the human annotators when comparing their ratings against the 'ground truth', in the form of the self-assessment by the subjects of the dataset. Three fragments were incorrectly classified as negative valence, whereas nine video fragments were misclassified as positive valence.

Discussion

We will now briefly discuss our results and highlight some areas for future work. First we have seen that our implementation of the method by [5] obtained a slightly higher performance than the original. Furthermore, Table 1 showed that the performance obtained by our human annotators was higher than that obtained by the automatic methods. We might even consider our automatic classifier advantaged as it was trained using other behavior shown by the same subject, as was also the case in [5]. Conversely it can be argued that this small amount of knowledge about the subject may not level the playing field when compared with human annotators with years of experience. Preliminary results indicate that the accuracy does not degrade when the automatic classifier is only trained using information from other subjects, suggesting that the knowledge of one specific subject might not be as advantageous as might be expected.

In our human annotation experiment, we found that the agreement among our raters is high with only 10 out of 48 video fragments receiving less than 75% of the ratings in one specific category. In Figure 2, we saw that three video fragments showing a positive valence were misclassified as negative valence. This effect might be caused by the subjects in these sessions showing little emotion, which makes it difficult for the observer to determine the experienced valence. We also found a significant fraction of negative valence video fragments which were misclassified by human observers. Similarly, in Figure 1, we see that many of the fragments misclassified by human observers were also mislabeled by the machine learning classifier.

Investigation of the misclassified sessions showed that several subjects show either little emotion or show expressions that might be difficult to relate to their self-assessment ratings. An example of the latter is a smiling response when confronted with a sad video. This difference between the observed emotions and the emotions that were reported by the subjects might be one important factor in explaining our results. The cause of the discrepancy could be because they actually felt the reported emotion (but did not clearly show this) or for example because they have given socially desirable answers.

In other comparisons between human and automatic annotations, the used datasets also differ. For instance, in [4], the subjects were asked to recall and describe a situation in which they experienced a specific emotion. There might be a difference in the expressiveness of the facial expressions depending on the type of emotion elicitation. Finally, where many similar studies provided the subjects with monetary rewards, we were not able to do so, which might have influenced the performance of our subjects.

Conclusion

In this paper we present the results of a comparison between classification accuracy of humans and machine-learning classifiers. For this we used the MAHNOB-HCI affective computing dataset and we have reproduced and extended the facial expression-based method by Koelstra and Patras. Our results show that both humans and machine classifiers agree to a large portion on the appropriate class for video fragments. In our experiments, we found that human annotators obtained higher accuracy than the automatic classification methods.

Future work could include evaluating how well the automatic classification method performs when training using the human labels, as described in [12], allowing for an estimation as to how close the performance of the automatic classifier is to the human annotators, rather than to the self-assessment ratings. Evaluating multiple datasets using the same human annotation method will also allow for comparisons across datasets, which could help compare automatic classification methods developed and tested using different datasets and allows datasets to be compared based on an human difficulty level of classification.

References

[1] Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, ***3**(1)*, 42-55.

[2] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T. ..., Patras, I. (2012). DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Transactions on Affective Computing*, ***3**(1)*, 18-31.

[3] Sneddon, I., McRorie, M., McKeown, G., & Hanratty, J. (2012). The Belfast Induced Natural Emotion Database. *IEEE Transactions on Affective Computing*, ***3**(1)*, 32-41.

[4] Janssen, J.H., Tacken, P., Vries, J.J.G. de, Broek, E.L. van den, Westerink, J.H.D.M., Haselager, P., IJselstein, W.A. (2013). Machines outperform laypersons in recognizing emotions elicited by autobiographical recollection. *Human–Computer Interaction* ***28**(6)*, 479-517.

[5] Koelstra, S., Patras, I. (2013). Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing*, ***31**(2)*, 164-174.

[6] Ekman, P, Friesen, W.V. (1977). Facial Action Coding System.

[7] Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, ***33**(1)*, 159–174.

[8] Den Uyl, M. J., Van Kuilenburg, H. (2005, August). The FaceReader: Online facial expression recognition. In *Proceedings of Measuring Behavior*, ***30.***

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. ..., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Jrnl of Mch. Learning Resrch*, ***12***, 2825-2830.

[10] Bould, E., Morris, N., Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition and Emotion*, ***22***(8), 1569-1587.

[11] Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, ***30***, 875-887.

[12] Steidl, S., M. Levit, A. Batliner, E. Nöth, H. Niemann (2005). Of All Things the Measure Is Man: Automatic Classification of Emotions and Inter-Labeler Consistency. In ICASSP (1), pp. 317-320.

Fishualization: A group feedback display

John Schavemaker[1], Erik Boertjes[1], Saskia Koldijk[1,3], Leon Wiertz[2], Suzan Verberne[3], Maya Sappelli[1,3], and Rianne Kaptein[1]

[1]TNO, Delft, The Netherlands
[2] Noldus Information Technology, Wageningen, The Netherlands
[3] Radboud University Nijmegen, Nijmegen, The Netherlands
[1] john.schavemaker@tno.nl

Abstract

In this contribution[1] we present a novel psychological intervention that maps human computer activity to a group feedback device on the basis of a combination of various types of unobtrusive, low-level sensors. The goal is to enable employees to gain insights into their working habits, to reduce stress levels and increase productivity. The unique approach taken is a social feedback board (*fishualization*), which gives collective feedback to employees of an entire department to stimulate social interactions, group awareness and openness, which are all beneficial for well-being at work. Fishualization is set up with a reconfigurable and modular approach to visualization and reasoning components to allow future additions and improvements.

Introduction

In our connected workplaces it can be hard to work in a calm and focused way. Ruff [5] speaks of 'plugged in compulsion' as "the strong need to check mail and the internet to stay in touch", and 'hurry sickness' as "the belief that one must constantly rush to keep pace with time". Stress can either directly lead to illness through its physiological effects or indirectly, through maladaptive health behavior, like smoking, poor eating habits or lack of sleep [1]. Certainly, some amount of stress is not harmful and might even be beneficial to gain concentration and focus, but extended periods of stress can be dangerous for health, e.g. causing burn-out.

Recent TNO research (2013) indicates that out of the Dutch workforce of 7.4 million people, 1 million have burn-out complaints and that stress and workload are identified as the main reasons for at least 7% of reported sick leaves with an estimated cost of roughly 900 million euro per year.

The goal of the Fishualization system is to enable employees to gain insights into their working habits to reduce stress levels, prevent burnout and increase productivity. The social feedback board increases awareness and stimulates discussion between employees and their differences in ways of working, which contributes to reduced stress and burnout prevention. Besides the feedback board for the whole department, an extension is provided with a personal feedback tool that provides personal feedback on working behavior, based on computer logging and input of ssubjective experience (energy level).

Related work

The SWELL fishualization system bears some similarity in function with the Hello.Wall from Fraunhofer [4], see Figure 1 (left). However, the Hello.Wall is considered to be 'informative art' and more abstract and futuristic without an explicit goal or application. When compared with the Hello.Wall, SWELL fishualization does not sense the vicinity of people for personalized feedback but it shares the concept of visualizing working patterns in a friendly and anonymous way without judgment. Both approaches also share the idea of using a reconfigurable and modular approach: the Fishualization system can easily be extended with new sensors, reasoning and mapping between data/information and visualized fish behavior.

[1] This publication was supported by the Dutch national program COMMIT/.

Figure 1. (left) Hello.Wall from Fraunhofer, (right) SWELL fishualization.

System description

The Fishualization system includes basic sensing using key logging (uLog [9]), a data storage and messaging framework (CommonSense [10] or RabbitMQ [11], both frameworks are possible), central aggregation and analysis of sensor data, and visualization of data at a central place (like the coffee corner) as feedback strategy. See Figure 2 (left) for a system architecture overview. The primary sensor is the key-logging software uLog, installed locally at each individual’s PC, that captures key strokes, mouse movements and clicks together with information provided by the operating system: window titles, active applications, application switches, etc. Other sensors and sensor processing may add other modalities like heart rate, dominant facial expression (from FaceReader [12]), and for example e-mail sentiments [7].

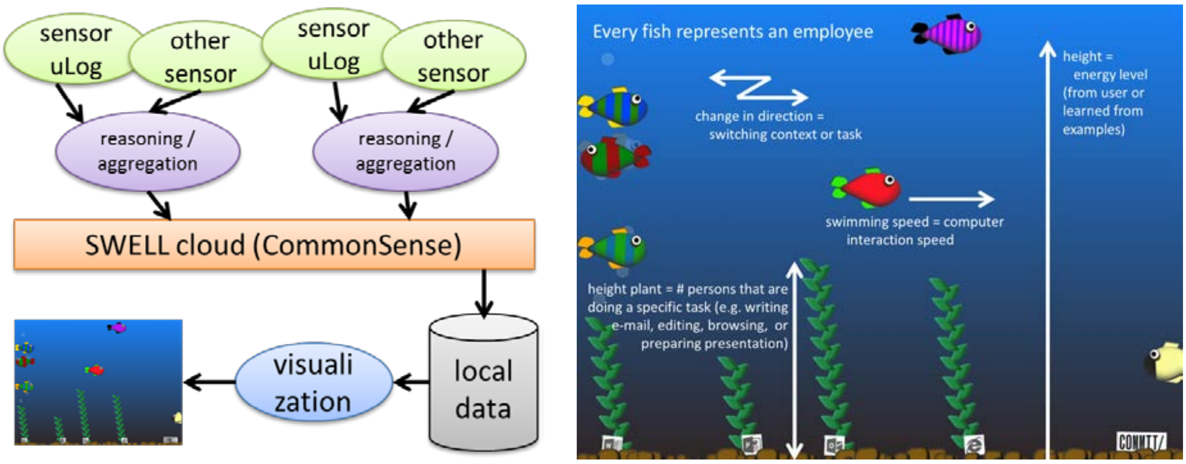


Figure 2. (left) Fishualization architecture (right) Fishualization legend.

The key-logging data is aggregated over time (1 minute) by the (local) reasoning/aggregation component. The aggregation keeps track of the keyboard and mouse activity as well as open applications and switches between them. Every minute the aggregation components post a data message to the SWELL cloud: CommonSense or a combination of RabbitMQ and an SQL database (in the experiments the RabbitMQ/SQL framework is used). The ‘fishualization’ shows visualizations of the state or mood of the group or individual people working in that specific group. Each fish represents an individual employee, see Figure 2 (right). The y-position of each fish represents the energy level of the corresponding employee on a scale of 0% (no energy) to 100% (full energy) and that is asked every 20 minutes by means of a pop-up dialog. The speed of horizontal movement of a fish is determined by how fast the corresponding employee is interacting with their computer (number of clicks and keystrokes) and the number of changes in direction per time unit represents the number of task or context switches per time unit. In this way a distinction can be made visible between fish (aka employees) that are in ‘the flow’ (no switches, high speed) and employees that work fragmented (many switches, low speed). Fish that leave the display at one side appear shortly thereafter at the opposite side with the same swimming direction. ‘Plants’ at the bottom of the screen represent active applications, for example, e-mail client, document editor, browser, or presentation editor. The more people work with a specific application, the larger the plant. In order to see your own fish behavior at the corner visualization works with a delay of seconds and an aggregation time of minutes.

Future additions will include real-time reasoning components to extract information to intelligently detect context switches (from one project to another) [6,8] and classify interaction patterns into pre-categorized task labels like ‘writing e-mail’, ‘editing document’, ‘browsing’, or ‘preparing presentation’, etc. [2,3]

Experiments

We are currently experimenting with the Fishualization group feedback device in our own working group with approval of management. A second and third deployment at other companies is anticipated in the near future.

Figure 3. Feedback device (PC and computer display) and camera and microphone for data collection.

The feedback device (a large computer display) has been placed in our coffee corner; a computer runs the web-based fishualization on the display, see Figure 3. In order to measure the effects of the deployment of Fishualization data collection was started 3 weeks before the official start. Using a webcam and a microphone, data was collected to measure activity at the coffee corner. Camera and microphone automatically deduce the number of detected faces, the amount of video motion and the average sound level. In order to warrant privacy of participants, images and sound are not stored but per minute aggregated values of number of detected faces, motion and sound volume are stored. The data collection is continued to compare activity statistics before and after deployment of fishualization. Figure 4 shows a typical day pattern at the coffee corner.

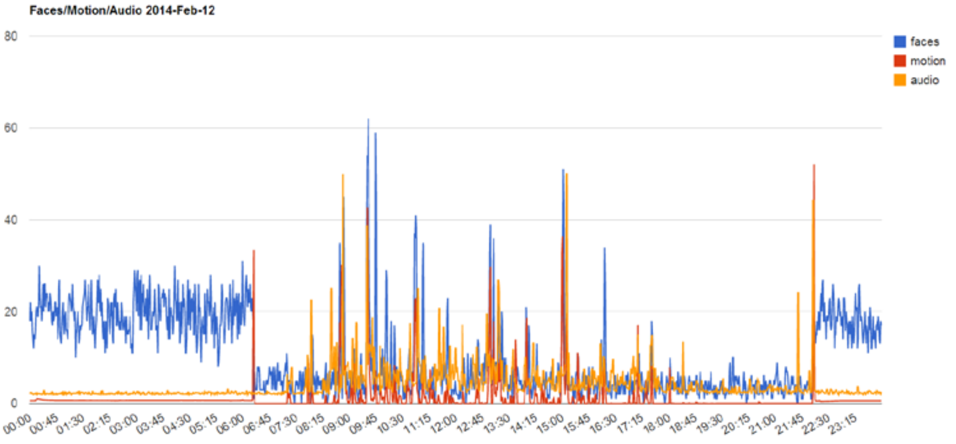


Figure 4. A typical day pattern at the coffee corner (the high number of detected faces between 22:30pm and 6:00am occur because the face detector is sensitive to noise when the lights are turned off in that period).

Evaluation

For future evaluation of the experiments we will use pre- and post-test questionnaires to measure several dependent variables that relate to the following claims:

C1: Collective department feedback stimulates social interactions, group awareness and openness, which are all beneficial for well-being at work.
C2: The feedback data is visualized in an intuitive, easily interpretable and appealing way so that it is most effective for gaining insights.
C3: The user's privacy is warranted, no detailed content information is shared (user's 'sensed' level of privacy is a key factor for acceptance of employees).
C4: Including a subjective variable (energy level) improves the personal awareness of well-being at work and its relations to working patterns.

Conclusions

In this paper we have introduced our fishualization concept. At the moment we are conducting experiments with this group display at one company and have plans for roll out at two other companies. Future additions will include reasoning methods for (semantic) interpretation of heterogeneous multi-scale sensor data providing information on context switches and task recognition (writing e-mail, editing document, browsing, or preparing presentation). The activity representations that are used can be easily extended with affect aspects as well; for example using analysis of facial expressions or analysis of e-mail sentiments (SWELL e-mail plugin).

References

[1] Bakker, J., Holenderski, L., Kocielnik, R., Pechenizkiy, M., & Sidorova, N.: Stess@ work: From measuring stress to its understanding, prediction and handling with personalized coaching. In: Proceedings of the 2nd ACM SIGHIT symposium on International health in-formatics, pp. 673-678. ACM (2012).

[2] Koldijk, S., Sappelli, M., Neerincx, M., & Kraaij, W. (2013). Unobtrusive monitoring of knowledge workers for stress self-regulation. In: *Proceedings of UMAP 2013* (Rome, Italy, 10-14 Juli 2013).

[3] Koldijk, S., van Staalduinen, M., Neerincx, M., & Kraaij, W. (2012). Real-time task recognition based on knowledge workers' computer activities. In: *Proceedings of ECCE 2012* (Edinburgh, Scotland, August 2012).

[4] Prante, T., Röcker, C., Streitz, N., Stenzel, R., Magerkurth, C., Van Alphen, D., & Plewe, D. (2003, October). Hello. wall–beyond ambient displays. In Adjunct Proceedings of Ubicomp (pp. 277-278).

[5] Ruff, J.: Information Overload: Causes, Symptoms and Solutions. Harvard Graduate School of Education, 1-13 (2002).

[6] Sappelli, M. The role of current working context in Professional Search. In: Proceedings of the 36th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2013), Dublin.

[7] Sappelli, M., Verberne, S., Kraaij, W. (2013) Combining textual and non-textual features for e-mail importance estimation. In: Proceedings of the 25th Benelux Conference on Artificial Intelligence (BNAIC 2013).

[8] Verberne, S., Sappelli, M., Kraaij, W. Term extraction for user profiling: evaluation by the user. Proceedings of the 21st Conference on User Modeling, Adaptation and Personalization (UMAP), Rome, Italy, 2013.

[9] http://www.noldus.com/human-behavior-research/products/ulog

[10] https://www.commonsense-dashboard.com/#/

[11] https://www.rabbitmq.com/

[12] http://www.noldus.com/human-behavior-research/products/facereader

An implicit behavioral approach to understand user satisfaction while preparing and consuming food at home

D. Labbe[1], A. Ferrage[1], N. Martin[1]

Nestlé Research Center, Lausanne, Switzerland

Abstract

We developed and tested a behavioral measure to better understand the entire experience and related satisfaction for a commercial packaged food. Measures were performed at home considering the entire food experience, i.e. food unpacking, food preparation and consumption. Participants filmed their interaction at home with the food and also scored in a questionnaire their satisfaction for a list of sensory and functional attributes as well as their overall satisfaction. Whereas quantitative analyses of behaviors provided insights explaining overall satisfaction, sensory and functional satisfaction rating did not. The implicit behavioral approach seems more adapted to understand user's satisfaction during interaction with a commercial packaged food at home compared to explicit questionnaire assessment.

Introduction

Traditionally, studies of consumer satisfaction with food have measured overall liking and/or specific sensory attribute liking through questionnaires at the moment of consumption [1]. They are usually administered in controlled environment (e.g. Central Location Test). Such practices fail to capture: 1) the entire experience related for instance to packaging opening or food preparation; and 2) the everyday context of consumption, which both strongly impact food satisfaction [2-4]. In addition, questionnaire assessment does not allow to capture processes that contribute unconsciously to food satisfaction such as the duration requested to open a packaging and the time needed for cooking. Behavioral observation could be a promising approach to apprehend the impact of such processes on consumer satisfaction. In this context, the objective of the present research was to assess the added value of implicit behavioral measure in a naturalistic environment, i.e. at home, to understand the impact of interaction with a commercial packaged food on user's satisfaction compare to outcomes from explicit questionnaire assessment.

Materials and methods

64 participants (aged between 20 to 55 years old) were recruited because of their willingness to buy and try the commercial packaged food chosen in this study. Participants were invited by groups of 6 to a 1-hour briefing session to receive the packaged food, the protocol instructions including a procedure to position the camera in their kitchen and to film themselves while interacting with the packaged food, a camera (Toshiba Camiléo P25, video resolution of 1920 x1080 pixel) and a tripod to stabilized the camera while filming.

Behavioral observations were based on videos taken by participants using a standard pocket camera to avoid the intrusive presence of a researcher. At home, participants set up the camera on its tripod so that their interaction with the packaged food was filmed from food unpacking stage to food preparation. After finishing the food preparation, participants rated their satisfaction for a list of 9 sensory (e.g. food appearance) and functional (e.g. food unpackaging) attributes covering their interaction with the food (the nature of the commercial food as well as the full list of attributes cannot be divulged for confidentiality purpose). This sensory/functional satisfaction rating is widely used in consumer research [5]. After food consumption, they scored their overall satisfaction for the entire experience, i.e. from food unpacking till the end of the consumption. Rating was performed using visual analogue scales anchored at the extremity with extremely unsatisfying and extremely satisfying.

A glossary combining 11 behaviors such as "food unpacking", "reading of food preparation instructions" (as for the list of attributes, the full list of behaviors is confidential) and definitions was developed during a pilot study performed in a kitchen of the Nestlé Research Center with 18 participants recruited from Lausanne and surroundings. The development of the coding scheme and coding of behaviors in terms of duration and/or frequency (for instance "food unpackaging" was only coded as duration) were performed in Observer XT 11.5 [6]. Trained staff from Noldus Information Technology company performed the behavioral coding.

Before conducting statistical analyses, behavioral and rating data non-normally distributed across participants were either log-transformed (for behavior durations and ratings) or squared root transformed (for behavior frequencies). A multiple regression analysis was computed on individual data to investigate the correlations between user overall satisfaction rating scores and the output variables, i.e. attribute rating scores and behavior frequencies/durations. Correlations were considered as significant when R-Squared p-value was below 0.05. All analyses were performed using NCSS software [7].

Results and discussion

The multiple regression model significantly estimated the overall satisfaction according to the output variables as R-Squared equal to 0.56 (p-value=0.01). Three behaviors (named A, B and C) out of the 11 significantly contributed to the overall satisfaction with R-Squared values equal to 0.08 (p-value<0.01), 0.07 (p-value= 0.01), and 0.04 (p-value=0.05), respectively. Sensory/functional satisfaction attributes measured with the questionnaire did not significantly contributed to the overall experience satisfaction as the R-Squared value for each of the 9 attributes did not reach a significant level. The final regression model was as follows:

*Overall satisfaction=-2.69*Behavior A frequency+1.68*Behavior B duration-1.24*Behavior C frequency+7.46*

According to this model, the longer the duration of behavior B and the lower the occurrence of behaviors A and C, the more satisfying the overall experience. The implicit behavioral approach highlighted that specific behaviors occurring during food unpacking and food preparation impacted the overall user's satisfaction but explicit questionnaire assessment did not significantly contributed to explain overall satisfaction.

Future perspectives

In the context of this study, the multiple regression modelling was an efficient tool to identify behaviors impacting consumer satisfaction while interacting with the product. Further research with different types of packaged foods is required to generalize the relationship between specific behaviors and consumer satisfaction and to validate the applicability of the multiple regression modelling. Today, it remains unknown whether behaviors occurring during the entire product experience can impact perception of specific food benefits (e.g. perceived quality, perceived healthiness) and additional investigations are required.

References

[1] Meilgaard, M., Civille, G.V., Carr, B.T. (1999). Affective test: Consumer tests and in-house panel acceptance test. In: *Sensory Evaluation Techniques* (3rd Ed), p. 221-235. CRC Press.

[2] Schifferstein, H.N.J., Fenko, A., Desmet, P.M.A., Labbe, D., Martin, N. (2013). Influence of package design on the dynamics of multisensory and emotional food experience. *Food Quality and Preference, **27***, 18-25.

[3] Meiselman, H. L. (2013). The future in sensory/consumer research:.....evolving to a better science. *Food Quality and Preference, **27***, 208-214.

[4] Piqueras-Fiszman, B., Spence, C. (2014). Colour, pleasantness, and consumption behaviour within a meal. *Appetite, **75***, 165-172.

[5] Moskowitz, H. R.; Krieger, B. (1995). The contribution of sensory liking to overall liking: An analysis of six food categories. *Food Quality and Preference, **6***, 83-90.

[6] The Observer XT software, < http://www.noldus.com/observer>, Accessed 28 February 2014.

[7] Hintze, J. (2001). Number Cruncher statistical systems, Karysville, Utah.

Objective measures of individual eating styles across different food and meal types

I. Ioakimidis, M. Tang-Bach, B. Langlet and P. Södersten

Department of Neurobiology, Care Sciences and Society, Division of Applied Neuroendocrinology, Karolinska Institute, Stockholm, Sweden.
ioannis.ioakimidis@ki.se

Introduction

Studying eating behaviour is important both in the fields of obesity and eating disorders [1,2]. However, while dietary habits receive a lot of attention [3], limited information exists on the behavioural patterns of individuals across meals at different times during the day or across meals with different food types. Does a person eat all their meals with a certain eating style? Is that style (partly) responsible for their weight and can we train the individual to normalize abnormal eating patterns? We believe that by using novel sensors and advanced data analysis we can answer these questions and integrate the generated knowledge into a novel, feedback-based guidance system. It is often the case that the quantification of eating behaviour is excluded from studies dealing with obesity or eating disorders, as it is widely believed to be the direct effect of various cognitive "processes" or "traits" (e.g., [4]). An unfortunate side effect of this assumption is that the collection of self-reported, cognitive data is regarded sufficient to describe eating, which is therefore not actually studied (for an example see [5]). Nevertheless, it has been recently noted that the actual eating behaviour should be focused upon, in eating disorders and obesity [1,2].

In the past, the objective quantification of the eating behaviour has led to the identification of two distinct eating styles (linear and decelerated) in normal individuals, based on the progression of eating rate as a meal progresses. The use of cumulative intake curves (CICs) to model single meals (i.e., quadratic equations fitted on intake measurements across the meal), has since led to numerous reports of specific eating styles in normal [6] and patient groups [7]. In summary, both the decelerated and the linear eating styles can be identified in normal individuals [8], while patient populations are characterised by linear eating [7]. Additionally, there is some evidence that an individual's eating style is relatively stable under similar conditions [8], but it can be modified by training [9]. This principle has been successfully used in the treatment of eating disorders [10] and obesity [11]. Other parameters have also been shown to affect the eating behaviours in humans [6]. One such factor is the type of the consumed food, which has a pronounced effect on the eating characteristics [12]. To our knowledge, there is no information about the conservation of the relative eating style of an individual across different food items.

In order to describe the behavioural elements that result in the recorded CICs, we videotaped meals recorded with the Mandometer. That allows us to add information on the actual chewing during a meal, in tandem with the CIC. In addition to chewing, video recordings are coded for the occurrence of spoonfuls (i.e., food removal from the plate) and bites (i.e., food entering the mouth). Using this data we developed a semi-automatic procedure for the correction of the recorded Mandometer data series, using the coded bites as objective anchor points. The validity of our correctional technique was assessed by comparison with two separate manual corrections performed by trained researchers [13]. The introduction of this semi-automatic procedure enables faster, systematic analysis of food intake and chewing, and it is ideal for behavioural comparisons across meals and individuals. In this case, we tested the hypothesis that humans can be characterized by their individual eating styles, irrespectively of the type of meal (e.g., breakfast, lunch or dinner) or the type of food that is consumed.

The analysis of eating behaviour across meals and food types for an individual will allow the calculation of comprehensive personalised eating profiles. In the future, similar information might be used to relate individual eating styles with risk behaviours that lead to obesity or eating disorders and can potentially be normalized by the use of specialized personalised guidance systems. Steps in that direction have already been made, with the launch of SPLENDID [14], an ongoing EU-project, that will mainstream, automate and integrate our methodologies of analysing eating behaviour in order to identify and modify behavioural patterns in real life.

Subjects & Methods

Subjects and meals. Fourteen healthy, normal-weight, female volunteers (age: 25.1±3.3 years, BMI: 22.7±2.9 kg/m^2; group characteristics are presented as mean value ± standard deviation), with no history of eating disorders, were offered lunch and dinner at two different occasions in our facilities (semi-controlled environment). In both cases, they were served ordinary Swedish food consisting of a mix of vegetables with chicken bits (henceforth referred to as *vegetables with chicken*: 426kJ, 10.7g protein, 8.2g carbohydrates and 2.5g fat/100g). In another occasion, twelve healthy, normal-weight women (age: 22.8±2.5 years, BMI: 21.9±1.6 kg/m^2) were served two different lunch types. In one occasion *vegetables with chicken* was served. In the other, the served food was a curry rise and chicken mix (*Nasi Goreng*: 598kJ, 5.5g protein, 18g carbohydrates and 5.7g fat/100g). Note that the two food types differ in their component composition, but they were selected to have fairly similar component sizes. Finally, thirteen comparable (age: 23.3±2.12 years, BMI: 22.5±2.54 kg/m^2) women were served either *vegetables with chicken*, or a pasta dish with minced meat and sauce (*macaroni*: 754kJ, 6.3g protein, 17.8g carbohydrates, 9g fat/ 100g), i.e., two foods which differ both in the nutritional composition and the component texture. In all cases the experimental sessions took place around 12.00 h and 18.00h, for lunch and dinner respectively. The repeated sessions per individual were randomized, separated by at least a week. In every case, subjects were initially familiarized with the experimental procedure during training lunches that were not analysed. All the participants were asked to follow their usual eating schedule during the experimental days outside the test meals. Since comparisons across conditions for the whole behavioural spectrum were planned, data was analysed jointly for decelerated and linear eaters. Also, only women were included in the presented datasets for the sake of simplicity.

Apparatus. Weight-loss data during the meals were measured using the Mandometer® (Mikrodidakt, Lund, Sweden) [7], a weighing scale linked to portable computer. The device records the reduction of the weight of the food on a plate placed on the scale with 1Hz during a meal. The meals were videotaped using a digital video camera (Digitalcam, Samsung, South Korea), positioned approximately two meters away and aimed at the plate and the maxillary-mandibular area of the participant.

Data collection. The Mandometer® data and the meal videos were transferred to a PC for further analysis. The occurrences of bites and chews were manually time-stamped on the video feeds (custom Excel macro). The video-generated data-series were automatically synchronized with the weight-loss data and the intake data series were automatically corrected as previously described [12]. The corrected food intake data were used to calculate the cumulative meal characteristics, i.e., the total food intake, and the meal duration. The CIC, a quadratic equation: $y = kx^2 + lx$, where y = food intake, k = rate of deceleration, l = initial speed of eating and x=time, was also calculated from the intake data series (Figure 1). The combined video and weight-loss data series were used to calculate the number and weight of bites, the number of chews, bursts of chewing and the pauses between bursts (data not presented). Data points across the meals were averaged over thirds of the meal for easier comparison. Additional anthropometric measures (e.g., % body composition; TANITA BC-418 MA, Tanita Inc., IL, USA) and subjective measurements around each meal (e.g., hunger and fullness) and the acceptability and the taste of the food (e.g., overall taste, smell, saltiness, sweetness etc.) were also collected. However, since they are not the main focus of the methodology they are not presented in detail here.

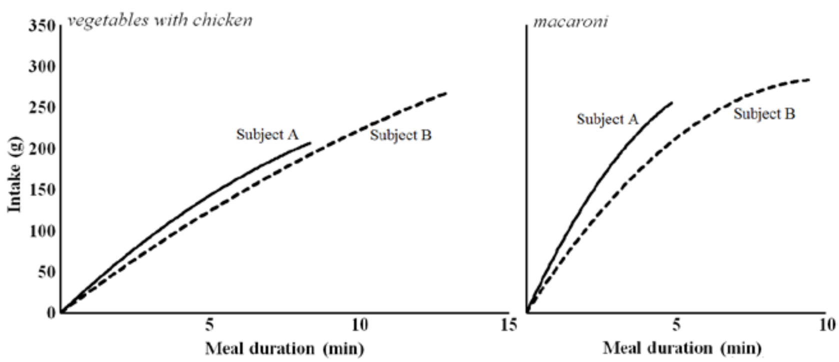


Figure 1. Data for two women being served different dishes for lunch. The presented eating curves have been fitted on weight-loss data collected by the Mandometer®.

Statistical analysis. All the statistical analyses were made using Sigmaplot 12.5 (Systat Software, CA, USA). Group differences were evaluated by using t-tests and analyses of variance, followed by post hoc tests, as needed. Correlations across conditions were performed in order to compare the relative similarities of the eating styles between different meals and food types.

Results

For sake of simplicity only results concerning the cumulative meal characteristics (i.e., total intake, meal duration, rate of deceleration and initial speed of eating) are presented here. This is the case for group comparisons, as well as measure correlations across conditions. The observed patterns mostly repeated (data not shown) when the detailed behavioural elements (e.g., chewing and bite rate progression over the meal etc.) were compared. Finally, group results are presented as means excluding measures of variability (for simplicity).

Group comparisons. Women ate about the same amount of food of *vegetables with chicken* at lunch and dinner (284 vs 275g respectively). The meals lasted about the same amount of time (9.5 vs 9.4min respectively). The eating styles across the meals (expressed through the CIC coefficients) were also not significantly different (coefficient k: -0.6 vs -0.7 and coefficient l: 35.5 vs 36.5 for lunch vs dinner, respectively). When women were served different foods with similar food itemization for lunch (*vegetables with chicken* vs *Nasi Goreng*), the participants ate similar amounts (282 vs 272g) in similar times (8.7 vs 9 min). The coefficients of the CIC were also similar (coefficient k: -0.7 vs -0.8 and coefficient l: 45.4 vs 42.4, respectively), showing that the different foods did not seriously affect the eating styles across the two food types. Finally, when women ate two very different foods for lunch (*vegetables with chicken* vs *macaroni*) they ate significantly different amounts (267 vs 312g, respectively, $p<0.01$). The *macaroni* dish was consumed faster (in 7.7 min vs 10.6 min for the *vegetables with chicken*) and the eating styles across the two food types were also significantly different (coefficient k: -0.8 vs -3.1 and coefficient l: 35.3 vs 63.2, for *vegetables with chicken* and *macaroni* respectively; $p<0.01$ in both cases). It is interesting to note that the subjects did not rate the foods differently in relation to their palatability ($p>0.1$).

Correlations. The total intake (r coefficient: 0.55, $p<0.01$), meal duration (r: 0.84, $p<0.01$), and the CIC characteristics (r: 0.4 and 0.39 for the coefficients k and l respectively, $p<0.05$ in both cases) correlated positively when women ate *vegetables with chicken*, either during lunch or at dinner, revealing that the relative eating styles of the subjects did not get affected by the type of the meal. When women consumed *vegetables with chicken* and *Nasi Goreng* for lunch, all the measures correlated strongly (data not shown) across the two food types, revealing that the relative eating styles across individuals remained unchanged. Finally, across the lunches with *vegetables with chicken* and *macaroni*, the eating styles remained relatively similar in all respects (see Table 1), even if the meal characteristics were significantly affected by the different food types (for an example see Figure 1).

Table 1. Correlations of measures across two lunches with different food types (vegetables with chicken and macaroni).

	Vegetables with chicken	Macaroni	Correlation coefficient r
Intake	267	312	0.78
Duration	10.6	7.7	0.94
Coefficient k	-0.8	-3.1	0.59
Coefficient l	35.3	63.2	0.83

Discussion

There is strong evidence that an individuals' eating style (i.e., linear eating) can be a risk of developing eating disorders or obesity [8, 10-12]. Here we present a methodology, based on objective measurement of eating behaviour in humans across different meals and food types, which can be a valuable tool for determining the baseline eating style of an individual. While the type of served food clearly affects a meal's characteristics (e.g., [12] and *vegetables with chicken* vs *macaroni*), individuals retain their relative personal eating style.

Additionally, the timing of the meal doesn't seem to affect the eating style when similar food is served (*lunch* vs *dinner*). Similarly, individuals eat with matching styles when foods with comparable mechanical properties are served (*vegetables with chicken* vs *Nasi Goreng*).

The presented methodologies can be further developed to allow detailed measurements over longer periods of time and across a wider range of meal types. More importantly, modern technology can facilitate the collection of such measures in real-life settings in wider populations. With those aims in mind, our methodologies have already been deployed by SPLENDID [14], an ongoing EU-project. We aim to develop a Personalised Guidance System for training children and young adults to improve their eating (and physical activity) behaviour, using eating behaviour during meals as an indicator for detecting users at risk for developing obesity or eating disorders.

Acknowledgements and statements

Part of the work leading to these results has received funding from the European Community's ICT Programme under Grant Agreement No. 610746, 01/10/2013 - 30/09/2016. All the authors have previously received/receive salaries by Mando Group AB, owner of the IP rights for the Mandometer. The protocols of the presented studies have been approved by the Regional Ethical Council in Stockholm, Sweden. We want to thank all the subjects for participating in our studies voluntarily.

References

[1] DiLeone, R.J., Taylor, J.R., Picciotto, M.R. (2012). The drive to eat: comparisons and distinctions between mechanisms of food reward and drug addiction. *Nature Neuroscience*, ***15***, 1330-1335.

[2] Walsh, B.T. (2011). The importance of eating behavior in eating disorders. *Physiology & Behavior*, ***104***, 525-529.

[3] Burkert, N.T., Muckenhuber, J., Großschädl, F., Rásky, E., Freidl, W. (2014). Nutrition and Health–The Association between Eating Behavior and Various Health Parameters: A Matched Sample Study. *PloS one*, ***9***, e88278.

[4] Finlayson, G., Cecil, J., Higgs, S., Hill, A., Hetherington, M. (2012). Susceptibility to weight gain. Eating behaviour traits and physical activity as predictors of weight gain during the first year of university. *Appetite*, ***58***, 1091-1098.

[5] Korinth, A., Schiess, S., Westenhoefer, J. (2010). Eating behaviour and eating disorders in students of nutrition sciences. *Public Health Nutrition*, ***13***, 32-37.

[6] Westerterp-Plantenga, M.S. (2000). Eating behavior in humans, characterized by cumulative food intake curves—a review. *Neuroscience & Biobehavioral Reviews*, ***24***, 239-248.

[7] Ioakimidis, I., Zandian, M., Bergh, C., Södersten, P. (2009). A method for the control of eating rate: a potential intervention in eating disorders. *Behavior Research Methods*, ***41***, 755-760.

[8] Zandian, M., Ioakimidis, I., Bergh, C., Brodin, U., Södersten, P. (2009). Decelerated and linear eaters: effect of eating rate on food intake and satiety. *Physiology & Behavior*, ***96***, 270-275.

[9] Zandian, M., Ioakimidis, I., Bergh, C., Södersten, P. (2009). Linear eaters turned decelerated: Reduction of a risk for disordered eating? *Physiology & Behavior*, ***96***, 518-521.

[10] Bergh, C., Callmar, M., Danemar, S., Hölcke, M., Isberg, S., Leon, M., Södersten, P. (2013). Effective treatment of eating disorders: Results at multiple sites. *Behavioral Neuroscience*, ***127***, 878-889.

[11] Ford, A.L., Bergh, C., Södersten, P., Sabin, M.A., Hollinghurst, S., Hunt, L.P., Shield, J. P. (2010). Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. *BMJ*, ***340***, b5388.

[12] Viskaal-van Dongen, M., Kok, F.J., de Graaf, C. (2011). Eating rate of commonly consumed foods promotes food and energy intake. *Appetite*, ***56***, 25-31.

[13] Ioakimidis, I., Zandian, M., Eriksson-Marklund, L., Bergh, C., Grigoriadis, A., Södersten, P. (2011). Description of chewing and food intake over the course of a meal. *Physiology & Behavior*, ***104***, 761-769.

[14] http://splendid-program.eu. Accessed 27 March 2014.

Fat Thoughts and Fat Content Do Not Affect Eating Behavior in Women

P. Södersten, M. Tang-Bach, M. Zandian, I. Ioakimidis and C. Bergh

Section of Applied Neuroendocrinology, Karolinska Institutet, Huddinge, Sweden
per.sodersten@ki.se

Abstract

The present study examined if the perception of fat and the content of fat in a food affect the cumulative curve of food intake and chewing among healthy women. One group of 11 women was served the same food twice, the food contained 2.5g fat/100g. They were informed verbally and by labeling that the food had a high fat content on one occasion and that it had a low fat content on the other. Another group of 13 women was also served the same food twice. In one condition the food contained 9g fat/100g and in the other it contained 3.1g fat/100g. The women were not informed about the fat content of the food. Cumulative food intake was measured using Mandometer® in both experiments and combined with video recording of the maxillary-mandibular region in the second, to measure the size of the bites, the sequences of chewing, and the pauses between bites. Labeling food "high fat" or serving food with a high fat content increased the perception of the fatness of the food but had no effect on food intake. Chewing was also unaffected by the fat content of the food. As a consequence, energy intake was markedly increased when the fat content of the food was increased. The results suggest that neither the thought of fatness nor the fat content of a food affect the cumulative intake of food and that energy intake can increase in the absence of a change in eating behavior.

Background

Women can be divided into those eating at a decelerating or constant speed over the course of a meal [1]. While decelerated eaters maintain their food intake, linear eaters lose control when experimentally challenged to eat quickly or slowly [2] and patients who have lost control, i.e., those who are under- or overweight, display a linear pattern of food intake [3]. Food intake is measured and experimentally manipulated using Mandometer®, a scale connected to a small computer, the subject removing food from a plate placed on the scale and adapting her/his speed of eating through visual feedback from the computer screen while eating [1]. Simultaneous video recording of the mandibular-maxillary region of the subject, thus synchronizing food intake with chewing behavior, has demonstrated that decelerated eaters take fewer and smaller bites by the end of the meal [4]. Because Mandometer® has been proven effective in treating both under- and overweight patients [5,6] and because the decelerated eater may be protected from losing control over body weight, the present study examined if this pattern of food intake and the associated pattern of chewing are affected by the information about the fat content of the food and the actual fat content of the food. Because the energy density of foods has been reported to enhance energy intake by increasing the speed of eating [7], it was hypothesized that an increase in the fat content of the food would increase food intake and the rate of deceleration and decrease the duration of the meal and the bite size by the end of the meal.

Methods

Participants: Healthy women, who were 22.3(2.3) years old and had a BMI of 22.1(2.1) kg/m^2, were recruited by advertisement on a nearly university campus.

Procedure: The women were informed about the project, ate a meal (mixed vegetables and grilled chicken cubes; 385kJ, 9.3g protein, 8.0g carbohydrates, 2.8g fat/100g, Findus AB, Bjuv, Sweden), using Mandometer® and the anatomy of their mouth region was video recorded in an introductory meeting.

One group of women (n=11) was served pasta with minced meat and a cheese sauce (427kJ, 6.1g protein, 13g carbohydrates, 2.5g fat/100g) on two occasions. On one occasion, the women were verbally informed that "the food has a high fat content", and on the other, they were informed that "the food has a low fat content" verbally and by labeling. Approximately 700g of food were presented in a lunch box, labeled either "high-fat" or "low-fat" in random order. The women served the food on to a plate and placed the plate on the scale of Mandometer®. They were encouraged to serve themselves and eat as much as they wanted.

Another group of women (n=13) ate macaroni with minced meat and a sauce that varied in fat content. Whipped cream (40% fat) was used in the high-fat condition (754kJ, 6.3g protein, 17.8g carbohydrates, 9g fat/ 100g), and milk (3% fat) was used in the low-fat condition (536kJ, 6.5g protein, 18.0g carbohydrates, 3.1g fat/100g). The women were not informed about the fat content of the food. The procedure was identical to that used in Experiment 1, but the meals were also video-taped.

The women were asked to estimate the fat content of the food in both experiments from “not at all” (score=0) to “extremely/very” (score=100) after eating the food.

Ethical statement: The procedures were approved by the Central Ethical Review Board of Stockholm.

Data analysis: Food intake was modeled by $y=kx^2+lx$; where y=food intake, k=rate of deceleration of the speed of eating and l=initial speed of eating. Results are show as box plots. Meals were divided into thirds and the number and weight of bites, the rate of chewing, the duration of bursts of chewing and pauses between bursts are shown as box plots as described before [8] and analyzed using two-way ANOVA for repeated measures and Tukey post-hoc tests (Sigmaplot 12, Systat Software, Inc. Point Richmond, CA).

Results

The perception of the fat content of the food increased to the same extent by labeling iso-energetic food as low or high in fat content or by increasing the fat content of the food, but neither procedure affected the cumulative curve of food intake (Figure 1). Simultaneous measurement of chewing showed that an increase in energy content of the food had no effect on eating behavior (Figure 2).

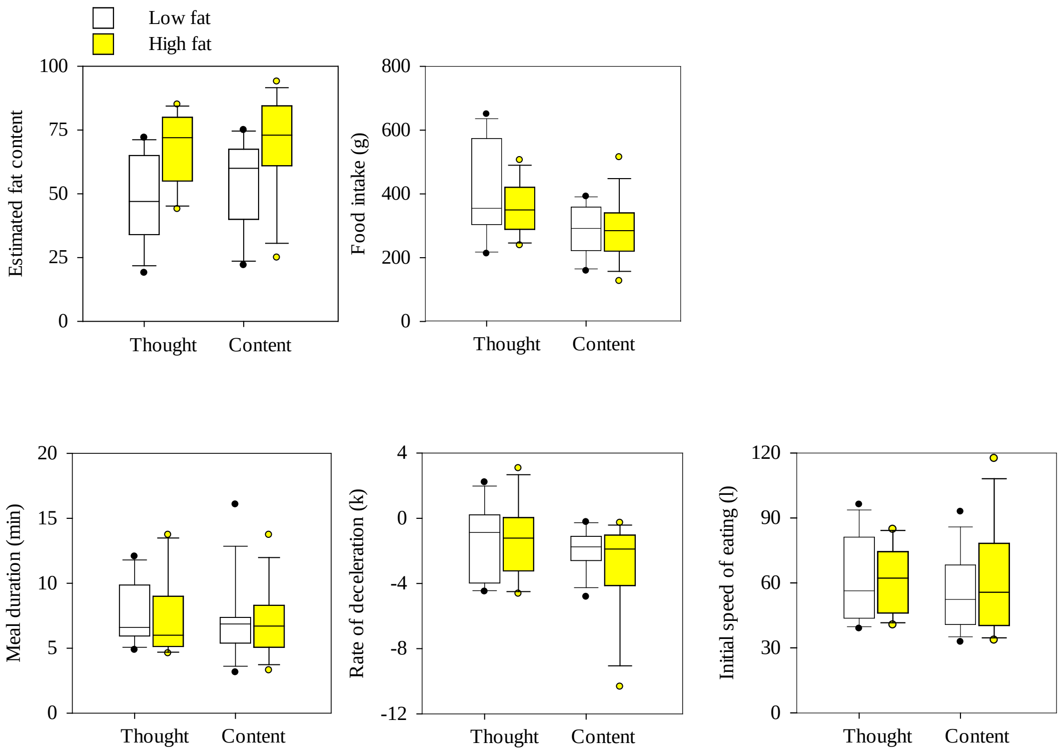


Figure 1. Estimation of fat content in a food labeled Low or High fat (Thought) or of a food with Low or High fat content (Content) and comulative food intake in women.

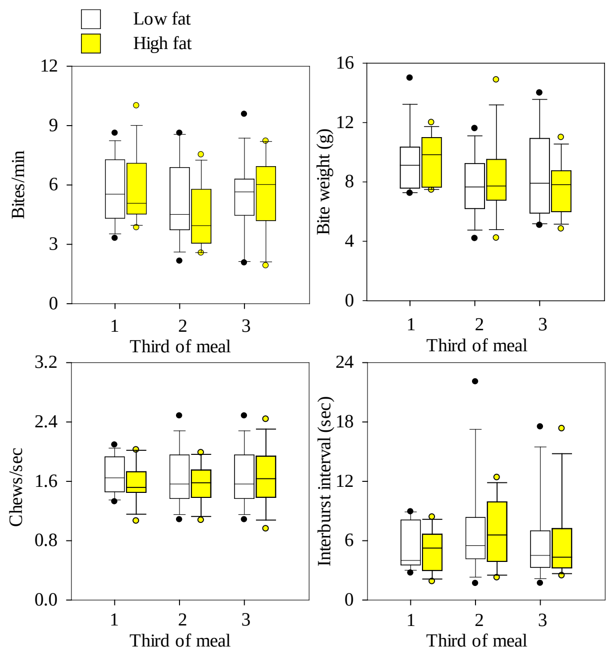


Figure 2. Bite size and weight, chewing and intervals between bursts of chewing in women.

Comment

While women, and probably men as well, can be induced to think that a food is low or high in fat content, although its fat content has not been changed, neither this maneuver nor a de facto increase in the fat content of a food significantly influenced their food intake. Also, eating behavior, i.e., chewing, was not affected by the fat content of the food. As a consequence, energy intake was markedly increased. Although the speed of eating was not determined by thoughts of fat food or by fat contents of food in the present study, it cannot be excluded that the differences in e.g., flavor and texture were too small to significantly affect eating behavior. Also, the subjects were healthy young women and they might not be as concerned with weight management to the extent that other groups of consumers are. These factors deserved attention in future experiment and other factors of potential interest include the physical characteristics of the food.

Acknowledgments

This work was supported by Mando Group AB.

References

[1] Zandian, M., Ioakimidis, I., Bergh, C., Brodin, U., Södersten, P. (2009). Decelerated and linear eaters: effect of eating rate on food intake and satiety. *Physiology & Behavior* **96**, 270-275.

[2] Zandian, M., Ioakimidis, I., Bergh, C., Södersten, P. (2009). Linear eaters turned decelerated: reduction of a risk for disordered eating? *Physiology & Behavior* **96**, 518-21.

[3] Ioakimidis, I., Zandian, M., Bergh, C., Södersten, P. (2009). A method for the control of eating rate: a potential intervention in eating disorders. *Behavior Reseach Methods* **41**, 755-760.

[4] Ioakimidis, I., Zandian, M., Ulbl, F., Ålund, C., Bergh, C., et al. (2012). Food intake and chewing in women. *Neurocomputing* **84**, 31-8.

[5] Bergh, C., Callmar, M., Danemar, S., Hölcke, M., Isberg, S., et al. (2013). Effective treatment of eating disorders: Results at multiple sites. *Behavioral Neuroscience* **127**, 878-889.

[6] Ford, A.L., Bergh, C., Södersten, P., Sabin, M.A., Hollinghurst, S. et al. (2010). Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. *BMJ* **340**, b5388.

[7] de Graaf, C. (2011). Why liquid energy results in overconsumption. *Proceedings of the Nutrition Society* **70**, 162-170.

[8] Ioakimidis, I., Zandian, M., Eriksson-Marklund, L., Bergh, C., Grigoriadis, A., et al. (2011). Description of chewing and food intake over the course of a meal. *Physiology & Behavior* **104,** 761-9.

The potential of the Intelligent Buffet in measuring food intake in a laboratory setting

Bent Egberg Mikkelsen[1] and Michal Dobroczynski[2]

[1]AAU-MENU, Research Group Meal Science & Public Health Nutrition Aalborg University
bemi@plan.aau.dk
[2]Syscore Aps

Introduction

Food behavior is increasingly studied experimentally in fields such as public health nutrition, health promotion and consumer science. At the same time new development in sensor technologies have made the measurement of the impact of such interventions easier. Use of automated intelligent devices for measuring and estimating food intake and behavior has been reported by various authors [3, 8 and 11]. One of the areas that has attracted particular interest is in the study of the dynamics of food choice in "ad libitum" settings such as buffets and cash cafeterias and a number of studies have explored the placement of food items and its influence on choice and food intake [6; 7, 9, 10]. Against this background a device that could make measurements of food intake in such environments easy, precise and reproducible under the name of the Intelligent Buffet (IB). The device was developed in the context of a research facility to study behavioral nutrition - the Foodscape Lab (www.foodscaeplab.aau.dk).

Aims & objectives

The aim of the paper is to contribute to the advancement of the science of automated measuring of food behavior using intelligent devices by proposing the Intelligent Buffet (IB). The specific objectives of the paper are:

- to present the features and abilities of the Intelligent Buffet (IB) for measuring food intake and to present examples of experiments that it can be used to measure.
- to give an account of the context in which the IB functions – namely the Foodscape Lab (www.foodscapelab.aau.dk) and to provide examples of the use of the IB.
- to present examples of the use of the IB and the types of questions that can been studied
- to discuss the potential of the IB technology in commercial out of home eating settings and its potential to allow for convenient day to day monitoring of dietary behaviour.

The development of the IB

In order to be able to artificially imitate the choice situation in which consumers would act at a buffet and to be able at the same time to apply a systematic research protocol that could measure automatically food taken in this situation an Intelligent Buffet was developed. The first prototype – the FoodScale Tracker - was developed in the framework of the graduate program Integrated Food Studies (www.ifs.aau.dk) in cooperation with company Syscore and the weighing scale supplier Mettler-Toledo [1].

The goal of the development of the device was to allow researchers to be able automatically to asses: "*who* is eating *what* in *which* amounts at *what* time" The IB was installed in the *serve area* of the lab (see Figure 1) and was developed to measure the impact of different types of food choice experiments on intake. The total set up involves a buffet, devices for identification of the individual test persons, scales to measure the amounts taken as well as overhead cameras connected to the Observer XT and that can be used as double check in play back mode to verify food choices of the subjects in the experiment. The IB was further developed from a commercial available buffet and subsequently fitted with sensors so that all test persons involved in the experiment could be accounted for when coding the data subsequently.

Figure 1. The Intelligent Buffet. To the left an overhead picture of the IB in the eat area of the FoodScape Lab. The left of the picture illustrates how the electronics is covered by the body of the buffet. The picture to the right shows the IB from the position of the overhead cameras. On the top right corner the picture shows inserted the wrist bands based on Near Field Communication technology that allows for distinguishing between the subjects taken food from the buffet. The wrist bands communicate with a receiver on the side of the buffet under each scale.

The scales were interfaced with a server to automatically detect "events" on the scale. This feature allows for setting up an experiment where all test persons are recognized through a Near Field Communication (NFC) wrist-tag. This functions the way that test persons will swipe the tag before taking food from the IB. The IB has 8 scales and allows for a buffet with a maximum of 8 different foods. After all identifications the server will detect the difference just before and just after the event in which food is taken from the buffet. The accuracy is +/- 2g

The IB technology is a setup utilising contactless technology, microcontrollers , open source electronics and a range of programming languages including C/C++, Python, PHP & bash for scripting. A typical IB experiment is setup as follows. First the protocol of the experiment is developed. This involve describing what kind of hypothesis the experiment should be testing. The development of the intervention/experimental idea can be guided either by theory or by pure speculation. Once the intervention has been defined the IB is set up and test persons are recruited. At the same time the video recording equipment is set up to record the full experiment. Test persons are registered and enter in the eat area of the lab where the IB is located. They will then do their food choices while their behaviour is detected by the IB controller software as well as recorded by the overhead video cameras. Apart from the requirement that the subjects have to swipe the chip before they take the food – there is nothing unusual and therefore minimum bias.

The FoodScape Lab

The IB is an integral part of the FoodScape Lab that offers a variety of technology for studying behavioural nutrition in a lab setting. The FSL is used in the study of the impact of choice architectures, nudges and other food environment innovations and are aimed at developing interventions that can be used for changing food and nutrition behavior in a healthier and more sustainable direction. The experiments in the FoodScape Lab serves as a pre-test that can allow for subsequent real life testing in living lab settings where buffets are applied for instance in canteens, schools, higher education, staff restaurants etc. The facilities are offered for evidence based research and education at Aalborg University, including Integrated Food Studies and PhD courses. The services offered are available for a broad range of research projects as well as for external users.

The Lab is divided into 3 experimental areas, one control and one analytical area (see Figure 2). The experimental areas include a *cook*, an *eat* and a *serve* area fitted with a rack system for placement of overhead cameras: The control room accommodates the serves and computers that operates overhead cameras. The analytical area (*analytics* room) is fitted with workstations that are used post experiment of processing and analysis of the data that is captured experimentally. The cameras are interfaced through a http protocol and as such uses existing standard cabling.

The analytics software includes the Observer XT software the Nvivo coding and analytical software that both allows for post experiment coding of video sequences captured on the overhead cams. The ArchGIS offers the possibility of storing GIS data on the location of food "opportunities" in the coal environment such as shops, restaurants, institutions within the welfare catering systems etc. The lab offers standard statistical software and includes SPSS, SAS; R & Stat and the Master cater software that allows for converting food intake into nutrients using the official Foodcomp database (www.foodcomp.dk) and to convert food intake into climate impact (carbon equivalents) using the Simapro LCA software. The analytics software in addition include custom built software to operate and analyse data from the IB and the Dietary Intake Monitoring System - DIMS that has been developed for automated recording and estimation of amount and type of food in- and outputs in out of home eating facilities such as hospitals [5].

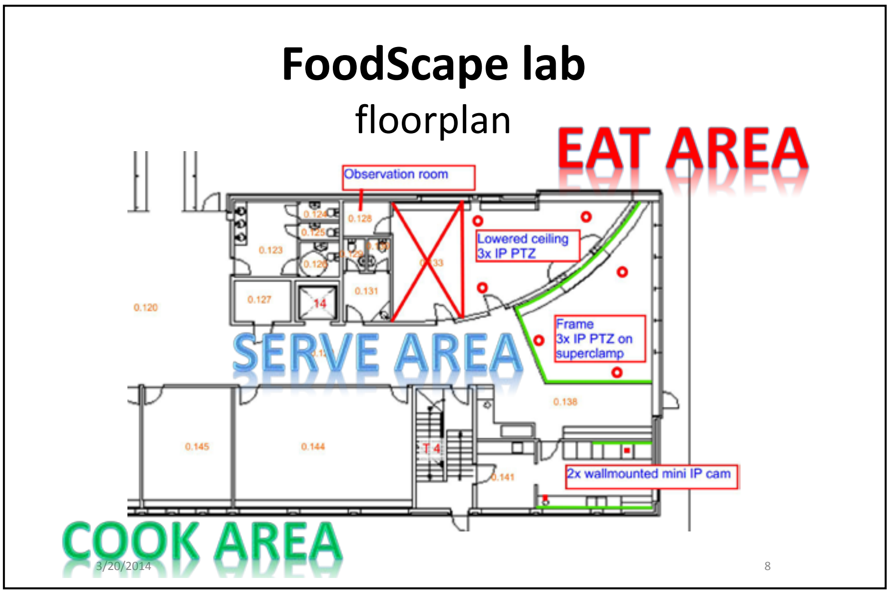


Figure 2. The FoodScape Lab. The figure shows the floorplan of the lab. The figure shows the division of the lab in an for cooking, an area for serving as well as an area for eating. All areas are fitted with a rack system allowing placement of overhead cameras. In addition to these three area the lab includes a control room that accommodates servers and computers. These are controlling the overhead cameras through joy-sticks. In the analytics room the software for post-experiment coding, analytics and other post-experiment activities are located

Sample projects using the IB

The development and fine-tuning of the equipment has been done in close cooperation with the supervisors at the Integrated Food Studies program in cooperation with the Syscore company. A few examples from the students use of the facility illustrates the kind of questions and experiments that can be approached with the IB.

The project *Nudging young men to eat more Fruit & Vegetables* [4] examined how choice architecture in a university canteen could be used to increase intake of fruit and vegetables single lunch serving among men aged 18- 29. The study aimed at determining the difference in the amount of self-served salad when increasing the amount of salad by 50 % and the accessibility of vegetables in trial compared to a control group. Although the study showed no significant difference it provided valuable insight in nature of buffet choice dynamics on efficient use of the IB.

In another study – the *Beat in the Music's Influence on the Food Intake* study [2] - it was investigated how the beat of background music would affect intake during a meal. A non participant observation approach was used for the experiment that was set up in the lab and that used campus students as the study subjects. The experiment investigated intake at no beat, down-beat and and up-beat conditions and was conducted using the Observer XT11 for observation of behavior and the PanelCheck V 1.4.0 for statistical analysis. From the study it was concluded that the beat in the music increased the food intake as well as the number of bites.

Acknowledgements

The Intelligent Buffet has been developed as a part of the foodscapelab in a cooperation between SysCore and Department of Development & Planning at Aalborg University, Copenhagen. Laurits Rohden Skov, Sofie Husby, Armando Perez-Cueto, Sanne Sansolios; FoodScape Lab and AAU. The first prototypes of the buffet were developed within the framework of the FoodServInspire project which is a part of the Inspire Food program financed by Agency for Research, Innovation and Education. Kwabena Titi Ofei; Bent Brandt. Erik Jellestad, Mettler Toledo; Michal Dobroczynski, Syscore; Henriette Hovmøller, and Ellen Bjerregård participated in this phase. FSL facilities are offered for evidence based research and education at Aalborg University, including Integrated Food Studies and PhD courses. The services offered are available for a broad range of research projects as well as for external users.

References

[1] Bjerregård, E. & Hertz, H. (2012) The FoodScaleTracker - Semi automatiseret metode til registrering af fødevareindtag ved buffet [in Danish]. (Semi automatiseed method for capturing food intake at a buffet), MENU Research Group, AAU

[2] Jensen, M.E.L., Granlund, S., Carstensen, R., Nielsen, V.H., Tingstedt, N., Hansen, M.V. (2012). *The Beat in the Music's Influence on the Food Intake* study. Student dissertation, University of Aalborg.

[3] McCabe-Sellers B. (2010). Advancing the art and science of dietary assessment through technology. *J Am Diet Assoc.*, ***110****(1)*, 52-54.

[4] Nielsen, B.N., Svane, D., Ahlmann, F.K., Schaldemose, H., Kongsbak, I.G., Atkinson, L., Wichmann, M. (2013). Nudging young men to eat more Fruit & Vegetables. Student dissertation, University of Aalborg,

[5] Ofei, K.T., Dobroczynski, M.T., Holst, M., Rasmussen, H.H., Mikkelsen, B.E. (2014). The Dietary Intake Monitoring System (DIMS) - an Innovative Device for Capturing Patient's Food Choice, Food Intake and Plate Waste in a Hospital Setting. Proceedings of Measuring Behavior 2014, 9th International Conference on Methods and Techniques in Behavioral Research (Wageningen, 27-29 August 2014) *[in press]*.

[6] Rozin, P., Scott, S., Dingley, M., Urbanek, J.K., Jiang, H., Kaltenbach, M. (2011). Nudge to nobesity I - Minor changes in accessibility decrease food intake. *Judgement and Decision Making*, ***6****(4)*, 323-332.

[7] Sigurdsson, V., Saevarsson, H., Foxall, G. (2009). Brand Placement and Consumer Choice: An In-Store Experiment. *Journal of Applied Behaviour Anaysis*, ***42****(3)*, 741-745.

[8] Thompson, F.E., Subar, A.F., Loria, C.M., Reedy, J.L., Baranowski, T. (2010). Need for technological innovation in dietary assessment. *J Am Diet Assoc.*, ***110****(1)*, 48.

[9] Van Kleef, E., Otten, K., van Trijp, H.C.M. (2012). Healthy Snacks at the Checkout Counter: A Lab and Field Study on the Impact of Shelf Arrangement and Assortment Structure on Consumer Choices. *BMC Public Health*, ***12****(1)*.

[10] Wansink, B. (2006) Mindless eating: Why we eat more than we think.

[11] Williamson D.A., Allen H.R., Martin, P.D., Alfonso, A.J., Gerald, B., Hunt, A. (2003). Comparison of digital photography of weighed and visual estimation of portion sizes. *J Am Diet Assoc.*, ***103****(9)*, 1139-1145.

The Dietary Intake Monitoring System (DIMS) - An innovative device for capturing patient's food choice, food intake and plate waste in a hospital setting

K.T. Ofei[1], M.T. Dobroczynski[2], M. Holst[3], H. H Rasmussen[3] and B. E. Mikkelsen[1]

[1]Research Group for Meal Science & Public Health Nutrition/Department of Planning & Development, Aalborg University, Copenhagen, Denmark
ofeikt@plan.aau.dk, bemi@plan.aau.dk
[2]SysCore ApS, Johan Krohnsvej 5, 3tv 2500 Valby Denmark
mic@syscore.dk
[3]Centre for Nutrition and Bowel Disease/Department of Gastroenterology, Aalborg University Hospital, Denmark
mette.holst@rn.dk, hhr@rn.dk

Abstract

Portable intelligent devices offer a number of opportunities to collect accurate dietary data for the convenient assessment of food choice and food intake. In this paper we describe a Dietary Intake Monitoring System (DIMS) as a device for capturing accurate data on a patient's meal both before and after consumption in a hospital setting for the assessment of food choice, food intake and plate waste. The DIMS consists of a digital camera, weighing scale, infrared thermometer, radio-frequency identification (RFID) reader, and a user RFID transponder card. The digital camera and weighing scale captures a digital photograph and measures the weight of content on a patient's plate before and after consumption. The temperature of food prior to consumption is also recorded with the aid of an infrared thermometer. Based on these inputs the DIMS gives an account of the amount of food on the plate as well as of the serving temperature of it. The DIMS prototype offers new opportunities to improve dietary data collection, routine monitoring of meal temperature and patients nutritional intake in hospital settings.

Keywords: Food choice, plate waste, food intake, hospital setting, RFID technology

Introduction

The nutritional well-being of hospital patients depends mainly on the nutrient intake from meals served by hospital foodservice. Undernutrition is a well-documented problem in hospitals around the world [1, 2]. A hospitals is a setting in which routine dietary monitoring can be of tremendous benefit and contribute positively to patient nutrition therapy, with a positive influence on foodservice quality and patient recovery [3,4]. Effective routine monitoring of patient food intake may depend on the availability of accurate dietary data, at the ward level. A major challenge is finding an appropriate method for the collection of accurate dietary data for large number of patients at nutritional risk, which can also reduce the workload involved in capturing the data. The method chosen is often influenced by the type of device, tool, or equipment to be used. For instance, when adopting a food photographic method for recording dietary data, the device required may vary from that of the weighing method. Although the weighing method is considered accurate, conducting it with the required tools and equipment for large numbers of patients can be very time consuming, costly and obstructive to the proceedings in a hospital setting [5].

The use of technological applications in dietary assessment methods have simplified the process and enhance the acquisition of data for the assessment of food choice, food intake and plate waste analysis [6, 7]. Technological innovations in digital cameras and communication devices that are able to collect dietary data in the form of digital images, have led to the development of a method for estimating food intake, that uses images of food choice and plate waste [8, 9]. It is possible to apply image process techniques or image analysis for recognizing food items from the digital photographic image estimate the weight of each food item and convert those input into calorie and nutrient values, from food composition databases [10, 11]. Estimating portion sizes based on images can induce errors, however incorporating the actual weight of a portion, has the potential to improve the accuracy of this approach [12]. This approach requires an innovative device that can capture and provide both serving size weights and digital photograph images to provide accurate dietary data. However, the availability of an innovative device that can combine both weighed and digital photography methods of collecting dietary data in a single device based on cutting edge technology is still in its infancy and the accuracy, convenience and feasibility of such a device is a considerable challenge. This gap in technology performance constitutes the background for the current study. Researchers, hospital and clinical nutrition professionals as well a Danish small- and medium-sized enterprise (SME) were involved. In this paper, we describe the Dietary Intake Monitoring System (DIMS) and its basic technologies, as well as the technique used for capturing accurate data of a patient's meal before and after consumption for assessing food choice, food intake and plate waste. In addition we present a pilot study, in which we tested the feasibility of the DIMS in a hospital setting for collecting and analyzing patient's food choice, intake and plate waste.

Description of key components of the DIMS

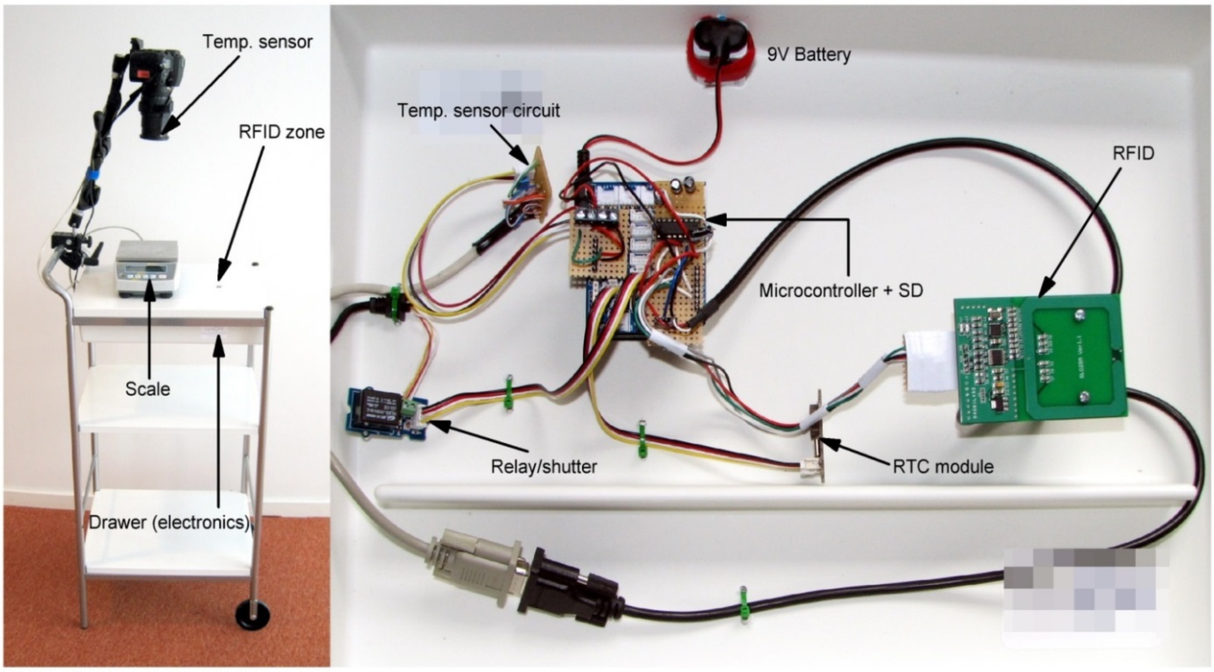


Figure 1. The DIMS prototype setup and components.

The DIMS prototype is built around an ATmega328 microcontroller that interfaces a shutter control relay, digital scale, radio-frequency identification (RFID) reader/writer module, infrared thermometer, real-time clock (RTC) module and a secured digital storage card for data collection. The heart of the system is contained in this single-chip microcontroller. The firmware for the microcontroller has been written in C computer programming language. The microcontroller and the sensors are powered by one 9V battery. It does not require external power lines; this increases DIMS's mobility and versatility.

Radio-Frequency Identification (RFID) transponder card

The RFID transponder card is coded with an identification number and links collected data to the patient. Since RFID numbers are unique, each patient is assigned one personal RFID transponder card. RFID is a contactless technology and can therefore also be used in situations where direct contact is not possible or not desirable. Many hospitals, other than those using traditional 1D and 2D barcode systems use wristbands with built-in RFID transponders. This makes the data collection process even easier although special care should be given to patients' privacy.

Digital Camera

A general purpose digital camera is used as the image capturing device for the DIMS setup. The aim of this special design is to minimize the amount of specialized equipment that has to be carried around when conducting on-site data collection. The camera is held on a special arm to allow for flexible positioning as shown in Figure 1. The camera is set to a high resolution in order to produce pictures that can enhance further analysis. In the DIMS setup the shutter has been configured to take approximately three pictures per plate over a period of 3 seconds. The shutter fires automatically once the weight on the scale has stabilized. Pictures are saved to camera's secured digital storage card and can be retrieved later for further analysis by the investigator.

Weighing Scale

The DIMS uses a certified trade digital scale (range 0-6000g, d = 1g) equipped with a RS232 data port so that the weight readout is automated. Weight data from the scale as well as other parameters are stored on DIMS's internal secured digital storage card. The digital scale takes on average 3 seconds to stabilize. Therefore all additional data acquisition processes are designed to be fast in order to keep the total time (per sample) very low.

Real-Time Clock (RTC)

The Real-Time Clock (RTC) component provides accurate time and date information for the system. It is also used as a key to data linking.

Infrared Thermometer

The infrared thermometer component of the DIMS was added since meal temperature is an important part of the patient's perception of meal quality and since sub-optimal temperature might be a food safety issue. It measures the average temperature of the food from a distance without having physical contact with the patient's food. The infrared thermometer face is directed towards the surface of the plate content, from which the average temperature of the food is determined by measuring the intensity of the infrared energy that accompanies the food. This feature can furthermore be used for routine meal quality assurance purposes.

DIMS Application

The DIMS application contains a simple piece of software that automates the process of data organization. It groups image files and joins them with the data collected from various sensors (RTC, RFID, and thermometer).

Measurement Procedures before and after Food Consumption for Plate Contents

The process for measurement has been designed with simplicity and speed in mind so that unnecessary delays in food serving are avoided. The measuring process can be described in four points:

RFID transponder card is placed onto the RFID reader active zone.

The plate is placed on the scale.

As soon as the weight stabilizes the system simultaneously takes pictures and collects other corresponding information on weight, temperature, date and time.

The plate and the RFID transponder can be removed. The system is ready to record another measurement.

Image data is stored on the camera's storage system whereas date, time, weight, temperature and RFID number are stored on the internal storage attached to the microcontroller of the DIMS. Once the data acquisition session is over, it is the investigator's responsibility to retrieve all the stored data onto a personal computer and run DIMS application.

Pilot test in a hospital setting

Methods

The DIMS was pilot tested in a medical gastroenterology ward. In this ward patients can choose their meals directly from the food trolley and be served by healthcare staff. We collected data at supper meal sessions for 3 days on three separate occasions. Our research contact person at the hospital identified potential patients for the study. The meals of twenty three participants were recorded in this study. Prior to the measurements we briefed the participants on the purpose of study and each was given a RFID transponder card. The principal investigator followed the 4 steps outlined in the measurement procedure for recording content on a patient's plate before and after consumption. The study was approved by the Local Scientific Ethics Committee and all patients gave oral informed consent.

Results and Discussion

The data was analyzed in the DIMS application, which created twenty three matched folders with photos of the content on a patient's plate before and after consumption. The DIMS application automatically generates a name for each folder by using patient's RFID code, date and time of measurement. We analyzed the each folder with the corresponding weight to deduce patient's food choice, plate waste, food intake and the food temperature before consumption as shown in table 2. In our pilot test we found that the DIMS is applicable for capturing photo images and measures the weight of content on a patient's plate before and after consumption in less than 4 seconds. The process does not obstruct meal servings at the ward. The process of matching before and after measurements, if done manually, will be tedious and subject to errors. However, with the DIMS application software, matching of before and after consumption data for each patient was less time consuming. Using the photo and weight data we were able to obtain vital information helpful to understand how much patients do throw away on the ward, food choice preference and intake. The goal of this paper was not to test DIMS for automatic estimation of energy and nutrient intake, since the DIMS application software is being developed to recognize food items from the image and estimate the weight. The data available from the DIMS indicates that the weight can improve the accuracy of estimating portion size from photo images for energy and nutrient intake calculation. This is because the total weight of all food items estimated from a photo can be equated to the actual total weight measured by the DIMS. To the best of our knowledge, no study has reported on the reliability and validity of a method that concurrently uses both photo and weight to estimate food choice, intake and plate waste. Therefore such research will be helpful to decide the appropriateness of using DIMS as method in a hospital setting.

Table 1. Shows food items chosen and served on the study days.

Meal Session	**Day 1**	**Day 2**	**Day 3**
Supper	Gullasch (GH) Mashed potatoes (MB) Parsley with red onion (PO) Potatoes small /spiced (PS)	Manogryde (MG) Fish (FH) Mashed potatoes (MP) Parboiled rice (PR) Hollandaise sauce (HS) Bacon tern (BT) Broccoli salad with pumpkin seeds (BP) Mayonnaise (MY) Butter (BU)	Chili concarne (CC) Mashed potatoes (MB) White sauce (WS) Carrot salad (CS) Fish cakes with herbs (FC) Potato / whole vaccum precooked (PP) Lingonberry jam (LJ) Butter (BU)

Conclusion

We have developed DIMS which is easy to use, and despite being in the prototype phase, requires less time to capture food choice, intake and plate waste in a hospital setting. This has been possible through a technology that integrates a digital camera, RFID sensor, weighing scale and infrared thermometer. The RFID technology provides more possibility to link data from the DIMS with other patients' information systems considered relevant for dietary assessment. Furthermore we are developing the DIMS application software for automatic recognition of the food items from the image, estimate the weight and convert into calorie and nutrient values from food composition databases. We believe that this technology can lead to a new methodological approach for collecting data for dietary studies and in particular to improve the monitoring of nutritional intake in hospitals.

Acknowledgements

The authors would like to acknowledge the support provided by the FoodServinSPIRe project (part of the InspireFood program) funded via Danish Agency for Science, Technology & Innovation and the Foodtura project funded via the Market Development Fund. The DIMS was developed in cooperation between researchers from Aalborg University, Aalborg University Hospital and Syscore ApS Company and builds on the technological insights from the FoodScape Lab (www.foodscapelab.aau.dk).

Table 2. Shows patient's food choice, plate waste, food intake and the temperature of food before consumption.

RFID code	Food items on plate before consumption (total food items)	Total food weight before consumption (g)	Temperature of food before consumption (C°)	Food items on plate after consumption (total food items)	Total food weight after consumption (g)	Plate waste (%)	Total food intake (g)	Total food intake (%)
a0789086	GH,MP, PO (3)	441	23,6	GH,MP,PO (3)	122	27,6	319	72,3
10a28e86	GH,MP,PO,AC (4)	358	24,0	(0)	0	0	358	100
90b59286	GH,MP,PO (3)	274	23,2	GH,PO (2)	54	19,7	220	80,3
80509386	GH,PO,PS (3)	379	23,1	PO,PS (2)	91	24,0	288	75,9
c0dc8e86	MG,PR,MY,BT,BP(5)	296	24,0	(0)	23	7,8	273	92,2
602b9086	FH,MP,HS,BT,BP(5)	430	23,3	(0)	0	0	430	100
c0279386	MG, PR,BP,MY(4)	202	23,8	MG,PR,BP,MY(4)	146	72,2	56	27,7
a0b49286	FH, PR,HS (3)	206	23,7	(0)	0	0	206	100
e0399386	FH, MP, HS,BT(4)	265	24,6	(0)	0	0	265	100
40ba9286	MG,PR,BP,BT,MY(5)	256	23,1	MG,PR,BP,BT,MY(5)	208	81,3	48	18,6
20f08f86	FH,MP,BU (3)	272	24,3	FH,MP(2)	157	57,7	115	42,3
a00d9186	FH,RP,HS,BT(4)	178	23,9	FH,RP,HS,BT(4)	160	89,9	18	10,1
e0be9286	MG,MP(2)	112	23,5	(0)	0	0	112	100
60b79286	MP,CC(2)	265	24,5	CC(1)	38	14,3	227	85,7
50319386	FC,MP,WS(3)	193	25,0	WS(1)	30	15,5	163	84,5
c04c9386	FC, MP,WS,CS(4)	270	26,2	(0)	0	0	270	100
d04c9386	FC, MP,WS,LJ(4)	302	25,7	(0)	0	0	302	100
80eb8f86	FC,MP,WS(3)	329	23,8	WS(1)	20	6,1	309	93,9
604e9286	FC, MP,WS,CS(4)	310	26,3	FC, MP,WS,CS(4)	244	78,7	66	21,3
504f9286	FC, MP,WS,CS,LJ(5)	264	24,3	FC, MP,WS,CS,LJ(5)	245	92,8	19	7,2
c0009386	FC,PP,WS,CS(4)	322	24,2	FC,PP,WS,CS(4)	117	36,3	205	63,7
20d98e86	FH,MP,CS,BU(4)	318	24,7	FH,MP,CS,BU(4)	132	41,5	186	58,5
904b9387	FC,MP,WS(3)	267	24,8	(0)	0	0	263	100
Mean	(4)	283	24,2	(2)	78	27,5	205	72,5

References

[1] Allison, S. P. (2003). Hospital food as treatment. *Clinical Nutrition*, ***22****(2)*, 113-114.

[2] Edington, J., Boorman, J., Durrant, E., Perkins, A., Giffin, C., James, R., Torrance, A. (2000). Prevalence of malnutrition on admission to four hospitals in England. *Clinical Nutrition*, ***19****(3)*, 191-195.

[3] Iff, S., Leuenberger, M., Rösch, S., Knecht, G., Tanner, B., Stanga, Z. (2008). Meeting the nutritional requirements of hospitalized patients: An interdisciplinary approach to hospital catering. *Clinical Nutrition*, ***27****(6)*, 800-805.

[4] Rasmussen, H. H., Kondrup, J., Staun, M., Ladefoged, K., Lindorff, K., Jørgensen, L. M., Wengler, A. (2006). A method for implementation of nutritional therapy in hospitals. *Clinical Nutrition*, ***25****(3)*, 515-523.

[5] Williams, P., Walton, K. (2011). Plate waste in hospitals and strategies for change. *E-SPEN, the European E-Journal of Clinical Nutrition and Metabolism*, ***6****(6)*, e235-e241.

[6] Ngo, J., Engelen, A., Molag, M., Roesle, J., García-Segovia, P., Serra-Majem, L. (2009). A review of the use of information and communication technologies for dietary assessment. *Br J Nutr*, ***101****(Suppl 2)*, S102-S112.

[7] Thompson, F. E., Subar, A. F., Loria, C. M., Reedy, J. L., Baranowski, T. (2010). Need for technological innovation in dietary assessment. *Journal of the American Dietetic Association*, ***110****(1)*, 48-51.

[8] Martin, C., Nicklas, T., Gunturk, B., Correa, J., Allen, H., Champagne, C. (2013). Measuring food intake with digital photography. *Journal of Human Nutrition and Dietetics*, ***27****(1)*, 72-81.

[9] Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C. J., Ebert, D. S., Delp, E. J. (2010). The use of mobile devices in aiding dietary assessment and evaluation. *Selected Topics in Signal Processing, IEEE Journal Of*, ***4****(4)*, 756-766.

[10] Islam, N. G., Dadabhoy, H., Gillum, A., Baranowski, J., Zimmerman, T., Subar, A. F., Baranowski, T. (2013). Digital food photography: Dietary surveillance and beyond. *Procedia Food Science*, ***2***, 122-128.

[11] Weiss, R., Stumbo, P. J., Divakaran, A. (2010). Automatic food documentation and volume computation using digital imaging and electronic transmission. *Journal of the American Dietetic Association*, ***110****(1)*, 42.

[12] Adamson, A. J., Baranowski, T. (2014). Developing technological solutions for dietary assessment in children and young people. *Journal of Human Nutrition and Dietetics*, ***27***, 1-4.

Measuring Food Behavior in the Restaurant of the Future

Stefanie Kremer and René de Wijk

WUR
stefanie.kremer@wur.nl

Most of the consumer decisions with regard to food choice and intake behavior are the result of sub-conscious rather than conscious processing. Sub-conscious processes are not readily accessible via introspection and self-report. Hence, we believe that food sensations, food choice and intake behavior can be best studied 1) with a combination of explicit (e.g., questionnaires) and implicit (e.g., physiological) tests in a (semi-) controlled laboratory environment, and 2) in a natural environment where consumer behavior can be observed and recorded objectively. For this purpose, the research facilities of the "Restaurant of the Future" was founded in 2007 in Wageningen, the Netherlands, as part of the Wageningen University Research Center. The research facilities consist of physiological, behavioral, and sensory laboratories and an actual instrumented lunch restaurant.

1) The laboratories consists of:
   - a chewing lab where oral movements during chewing are recorded in detail with a 3-D electromagnetic articulograph. Ongoing studies are aimed at identifying the relationships between specific types of oral behavior and specific sensations (e.g. [6,7]. These results can be used by food companies to optimize their foods with regard to desirable sensations.
   - a physiological lab where bodily responses, such as facial expressions (automatically registered and analyzed with FaceReader software), heart rate and skin conductance responses, to food-related stimuli are collected. These stimuli can involve the foods themselves but also their aromas, packaging, brand names, etc. Bodily reactions are probably indicative of product emotions, an important determinant of food choice and purchasing behavior (e.g. [2,4,8].
   - The mood rooms, where light, sound, and odor ambiences can be varied systematically and their effects on consumer behavior towards food and non-food stimuli can be assessed with objective and subjective tests (e.g. [4].
   - a sensory laboratory where food sensations are assessed using state of the art taste tests by both experts and naive consumers.
2) The lunch restaurant, consists of an actual instrumented lunch restaurant equipped with a series of ceiling-mounted video cameras (powered by Noldus Information Technology software), a built-in weighing scale, an automatic consumer tracking system, and automated cash registers. More than1000 registered visitors, all WUR employees, visit the restaurant regularly and sometimes on a daily basis. During their visits, food choices of each identified visitor are recorded automatically, providing insight in the composition of lunches and in determinants of repeated lunch selections of individual consumers, and allowing answers to questions such as: are repeated selections based on energy, or preference for specific types of products and nutrients, or other variables such as price and weight? (e.g. [1,3,9]. Other methods are used to achieve more in-depth insight in consumer processes at the moment that lunch selections are made. The cameras and tracking systems monitor the routing of the consumers between the various buffets providing information on frequency and duration of inspection and whether or not this results in a purchase (e.g.[5]. If required, selected visitors can be equipped with eye tracking systems and physiological sensors to provide information on intentional processes and product emotions.

References

[1] Bolhuis, D.P., Temme, E.H.M., Koeman, F., Noort, M.W.J., Kremer, S., Janssen, A.M. (2011). A salt reduction of 50% in bread does not decrease bread consumption or increase sodium intake by the choice of sandwich fillings. *Journal of Nutrition,* ***141***, 2249–255.

[2] De Wijk, R.A., Kooijman, V., Verhoeven, R., Holthuysen, N, de Graaf, C. (2012). Autonomic nervous system responses on and facial expressions to the sight, smell, and taste of liked and disliked foods. *Food Quality and Preference*, ***26***, 196–203.

[3] Kremer, S., Derks, J., Nijenhuis, M.A., Boer, E., Gorselink, M. (2012). Effect of a holistic meal and ambiance concept on main meal enjoyment and food intake of Dutch nursing home residents: A pilot study. *Journal of Aging Research & Clinical Practice*, ***1***, 237–244.

[4] de Wijk, R.A., Zijlstra, S. (2012). Differential effects of exposure to ambient vanilla and citrus aromas on mood, arousal and food choice. *Flavour****, 1***,24.

[5] Hinton, E.C., Brunstrom, J.M.' Rogers, Fay, S.H.,' de Wijk, R. (2013). Quantifying plate-cleaning using 'The Restaurant of the Future. *Appetite*, ***63***, 31–35.

[6] Bolhuis, D.P., Lakemond, C.M.M., de Wijk, R.A., Luning' P.A., de Graaf, K. (2013). Consumption with large sip sizes increases food intake and leads to underestimation of the amount consumed. *PLoS One,* ***8****(1),* e53228.

[7] Bolhuis, D.P., Lakemond, C.M.M., de Wijk, R.A., Luning' P.A., de Graaf, K. (2014). Both a higher number of sips and a oral transit time reduce ad libitum food intake. *Food Quality and Preference,* ***32****(Part C)*, 234–240.

[8] De Wijk, R.A., He, W., Mensink, M.M., Verhoeven, R., de Graaf, C. (2014). ANS responses and facial expressions differentiate between repeated exposures of commercial breakfast drinks. Accepted for publication in *PLoS One,* ***9****(4),* e93823.

[9] Janssen, A.M., Kremer, S., van Stipriaan, W.L.,Noort, M.W.J, de Vries, J.H.M., Temme, E.H.M. (s*ubmitted*). Reduced-sodium lunches are well accepted by uninformed consumers over a three-week period and result in decreased daily dietary sodium intakes.

How to use a living lab restaurant to study satiety and food intake in a real meal situation?

Agnès Giboreau

The Center for Food and Hospitality Research, Institut Paul Bocuse
agnes.giboreau@institutpaulbocuse.com

Introduction

Studying consumers' preference leads to raise the question of judgment variability, linked to the sensory quality of the products but also to subject differences and context influence. Context effects are more and more described such as interpersonal conditions, product presentation, label, price, physical environment etc. [e.g. 7, 2, 4, 6 & 15]. However, many food studies have grown on a *Stimulus – Response* model but such laboratory methods to get knowledge on food products do not take contextual factors into account.

Thus, it is clearly needed to consider food perception in a synthetic perspective determined by the object, the subject and the eating context [e.g. 12, 13]. Considering food preference in this triad perspective, new methodologies are developed, based on choices and decision making theories, on usages and practices descriptions, on perception and cognition models. Real-life situation approaches allow pointing out the complexity of food behaviour as opposed to controlled laboratory stimulus based approaches.

More specifically, experimental environments mastering *the food, the eater and the meal environment* are now available through experimental restaurants where cooks control the recipe, where consumers are clients and where the ambiance is also controlled (http://www.institutpaulbocuse.com/us/food-hospitality/).

Principle

The multidisciplinary environment of the Institut Paul Bocuse Living Lab offers a place where scientists work together with chefs and headwaiters, sharing points of view, methodological feedback and research results from their respective fields of competences.

The meal offer, the eaters and the environment are designed or selected depending on the objective of each study, reproducing the real-life situation under study in the Living Lab. Several studies will be presented to illustrate the diversity of possible designs:

- A school restaurant providing school meals and inviting children for their lunch break on a regular school day [8]
- A workplace restaurant, providing a single menu offer and gathering colleagues at the same table for a free meal [3]
- A cafeteria restaurant with a self-service device available for any consumer at a reasonable price [9, 14]
- A gastronomical restaurant with a 2 starters / 4 main courses offer with paying customers coming for dinner in groups of relatives – friends or family [10, 11]

More controlled buffet designs are also possible to study food intake and satiety in a naturalistic environment at the behavioural, psychological and biological levels [1].

Environment: The Living Lab is designed to replicate actual cooking and consumption situations. Both the floor and the ceiling of the restaurant and the kitchen are modular to adapt to different equipment configurations, allowing for the construction of different eating contexts. We are thus able to create every type of restaurant environment: traditional, collective, brewery, cafeterias, etc. as well as to change the kitchen configuration for specific studies. This includes changes in the ambiance and décor with adjustable temperature, sound system, visual environment, lighting, table dressing, plates and cutlery etc.

Participants: Users are key components of each Living Lab study. Chefs, waiters, consumers are involved as actors of real-life situations – captured through in situ methods. Users are not necessarily aware of the whole range of issue of the study; discrete data acquisition material and adequate contextual setting provide a strong sense of real-life situation to users. This is particularly true for the experimental restaurant when it is used as a commercial restaurant.

Food: The food is prepared by students in culinary art programs under the supervision of a chef instructor. The selection of raw material, ingredients, and processes is made on the basis of the real-life situation to reproduce, both qualitatively and quantitatively, and in respect with supply chain and cost constraints. The way food is displayed, served, sold is also defined in line with the natural reference.

Data collection: The Living Lab is equipped with a data acquisition system (video cameras, microphones, and computerized questionnaires) as well as a laboratory for biomedical sampling and food control analyses. Each work is based on human behaviors through observation, questionnaires or experimental tasks. Ethnographic studies are done through video, interviews and food preference through questionnaire (visual analog scales), experimental economics (willingness to pay) or behavioral data (choices, served or consumed food). With the collaboration of academic partners (CENS, The European Center for Nutrition and Health, CRNL, The Lyon Neuroscience Research Center), complementary physiological and nutritional analyses can be performed and coupled with behavioral analysis. The combination of all tools allow the acquisition and the comparison of behavioral, psychological and physiological date

Conclusion

To sum-up, the Living Lab experimental Restaurant allows the Center for food and Hospitality Research to develop an original research strategy that takes into account three major concerns:

- The taste and pleasure of a meal, tasting, table etiquette, service, hosting and hospitality;
- Health and well-being, nutritional balance, physiological and mental well-being;
- The environment of people eating, in its economic, social and contextual aspects.

References

[1] Allirot X, Saulais L, Disse E, Nazare JA, Cazal C, Laville M. (2014). Integrating behavioral measurements in physiological approaches of satiety. *Food Quality and Preference, **31***, 181-189.

[2] Cardello, A.V. et al. (2000). Predictors of food acceptance, consumption and satisfaction in specific eating situations. *Food Quality and Preference, **11**(3)*, 201-216.

[3] Fernandez, P., Bensafi, M., Rouby C., Giboreau, A. (2013). Does olfactory specific satiety take place in a natural setting? *Appetite, **60**(1)*, 1-4

[4] Hersleth, M. et al. (2003). Effect of contextual factors on liking for wine Çöuse of robust design methodology. *Food Quality and Preference, **14**(7),* 615-22.

[5] Jacquier C., Giboreau A. (2012). Environmental influence on emotional state related to food liking at the restaurant. *Proceedings 5th European Conference on Sensory and Consumer Research*, 9-12 September, Bern, Switzerland.

[6] King et al. (2000). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food Quality and Preference, 15(7)*, 645-653.

[7] Meiselman, H.L. et al. (2000). Demonstrations of the influence of the eating environment on food acceptance. *Appetite, **35**(3)*, 231-237.

[8] Morizet D., Depezay L., Combris P., Picard D., Giboreau A. (2012). *Effect of labeling on new vegetable dish acceptance in preadolescent children. Appetite, **59**(2)*, 399-402.

[9] Petit E., Saulais L., Rouby C., Giboreau A. (2011). How much do you like this dessert? Role of nutritional information in the restaurant on consumer hedonic and economic preferences. *9th Pangborn Sensory Science Symposium*, 4-8 September, Toronto, Canada.

[10] Piqueras-Fiszman, B., Giboreau, A., Spence, C. (2013). Assessing the influence of the color of the plate on the perception of a complex food in a restaurant setting. *Flavour*, ***2****(24)*, 1-11.

[11] Porcherot C., Petit E., Giboreau A., Gaudreau N., Cayeux I. *Measurement of self-reported affective feelings when an aperitif is consumed in an ecological setting.* Submitted FQAP.

[12] Renner, B. et al. (2012). Why we eat what we eat: The Eating Motivation Survey (TEMS). *Appetite,* ***59****(1)*, 117-128.

[13] Robinson et al. (2013). The influence of recent tasting experience on expected liking for foods. *Food Quality and Preference* ***27****(1)*, 101-106.

[14] Saulais, L. (2013). Using Nudges to guide food choices in the restaurant: an experimental evaluation of two nutritional interventions. *Proceedings ICCAS 2013 International Conference on Culinary Arts and Sciences*, 19-21st June 2013, Porto Portugal.

[15] Sester, C. et al. (2013). Having a drink in a bar: An immersive approach to explore the effects of context on drink choice. *Food Quality and Preference*, ***28****(1)*, 23-31.

Home cage-based long-term monitoring of fear in mice: Novel approach to determine individual differences in risk assessment and avoidance

O. Stiedl[1,2], T. Hager[1,3], R.F. Jansen[1], A.W. Pieneman[1], S.N. Mannivanan[4], I. Golani[5], A.B. Smit[2], M. Verhage[1]

[1]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), VU University Amsterdam, The Netherlands
oliver.stiedl@cncr.vu.nl
[2]Department of Molecular and Cellular Neurobiology, CNCR, VU University Amsterdam, The Netherlands
[3]Sylics B.V., Amsterdam, The Netherlands
[4]Biobserve GmbH, St. Augustin, Germany
[5]Department of Zoology, Faculty of Life Sciences and Sagol School for Neuroscience, Tel Aviv University, Israel

Introduction

Behavioral phenotyping is indispensible for biomedical research. Fear conditioning has been one of the most successful behavior assays in behavioral neuroscience, to investigate the neural systems and molecular basis of various aspects of emotional learning across a wide range of species [1]. Dysfunction of the fear circuits is assumed to underlie mechanisms of affective disorders and is frequently investigated in rodent models. However, many currently used behavior assays, including classical fear learning tests are geared for higher throughput. Thus, they have short test duration and require frequent human interference, inducing considerable data variation, interpretational ambiguity and limited translational value. Except for few studies employing novel approaches [e.g. 2], there is a general lack of progress in behavioral neuroscience for measuring and interpreting behavioral responses of laboratory animals particularly under ethologically valid conditions [3].

Material and Methods

In view of the above-mentioned limitations, we developed a fully automated home cage system (DualCage; see *HomeCage*[Plus], Biobserve, St. Augustin, Germany) consisting of a safe home compartment (HC) coupled to a risk-prone test compartment (TC) separated by a controllable door as in passive avoidance experiments [4]. This approach allowed the investigation of conditioned contextual fear responses and concomitant behavioral changes in mice from baseline behavior including circadian activity, via fear acquisition, consolidation with post-shock activity assessment, and retention to extinction determined by TC re-exploration. Behavioral responses were based on deliberate choice, motivated exclusively by novelty-seeking and were monitored over several days without human intervention. This approach may help to avoid the negative consequences of unspecific stressors on behavioral responses. A separation between HC and TC is necessary for exploiting novelty-seeking behavior and decision-making at distinct times. The resolution of the system is demonstrated by comparing the performance of male mice of the two closely related substrains C57BL/6J (6J) and C57BL/6N (6N). Compared to 6J mice, 6N mice show stronger fear responses in fear conditioning followed by delayed extinction [e.g. 5], which was reconfirmed by us in classical fear conditioning. When a mouse had fully entered the TC, the door was closed for inescapable shock exposure. Mice of different groups were subjected to various levels of negative reinforcement based on the number (0, 1, 3 or 5) of 2-s foot shocks (0.7 mA, scrambled) presented 30 s after their first TC visit to determine the impact of reinforcement levels on retention performance. After the training session, the door was opened and mice returned to their HC. Different levels of reinforcement were used because classical fear conditioning studies based on freezing reported an increase of fear suggesting a simple form of learning, whereas complex forms of learning show an impairment with increasing arousal ("stress") as indicated by the Yerkes-Dodson law [reviewed in 6]. Since passive avoidance and contextual fear conditioning are both hippocampus-dependent learning tasks [4,7], the lack of fear learning impairment in classical fear conditioning by higher levels of reinforcement is surprising [6]. Twenty-four hours after training, contextual fear memory and its extinction were assessed for 48 h based on deliberate TC revisits and re-exploration. This included the analysis of the body posture of mice based on 3-point tracking to assess stretch-attend postures as an unambiguous sign of risk assessment in the face of threat [8]. The study was approved by the animal research committee of the VU University Amsterdam and conducted according to Dutch regulations in compliance with the European Council Directive (86/609/EEC).

Results and Discussion

Comparison of circadian activity between the two mouse substrains identifies a significant difference in locomotor activity that went unnoticed in classical tests [e.g., 5]. 6N mice showed an increased latency to re-enter the TC suggesting increased fear compared to 6J mice. These long retention transfer latencies (several hours) indicate long-lasting fear response, and thus emphasize the importance of extended time scales for behavioral monitoring. TC visits were predominantly confined to the dark phase. Considerable inter-individual performance variation within both isogenic substrains suggests epigenetic contributions to individual differences and highlight different, i.e. active or passive, coping styles [9]. Sub-populations of mice that never re-entered the TC constitute a clear model of post-traumatic stress disorder. In contrast to classical fear conditioning and passive avoidance results, there was no difference in fear extinction in 6J versus 6N mice that re-entered the TC. This suggests that unspecific arousal ("stress") in classical behavior tests impairs extinction learning in 6N mice, since extinction of conditioned heart rate responses in the home cage under conditions of reduced unspecific "stress" also failed to show differences between the two substrains [5].

Conclusions

This novel approach improves ethological and face validity while reducing interpretational ambiguity of emotional states in models of fear learning, memory and anxiety-related disorders. This is accomplished by exploiting automated behavioral measures of high translational value such as risk-assessment and avoidance - core endophenotypes of affective disorders. This novel approach is thus a highly valuable asset for extending and rectifying interpretations based on classical fear behavior tests. Our approach is currently being extended to other mouse strains including mutants and pharmacological intervention studies, for further validation.

Acknowledgements

We thank Dr. Christian Gutzen and colleagues from Biobserve for thorough hard- and software support. T.H. and S.N.M. were supported as early stage researchers by the European Union Seventh Framework Programs under grant agreements no. PEOPLE-ITN-2008-238055 (BrainTrain) provided to O.S. and Dr. Christian Gutzen.

References

[1] LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Reviews of Neuroscience,* ***23***, 155-184.

[2] Fonio, E., Benjamini, Y., Golani, I. (2009). Freedom of movement and the stability of its unfolding in free exploration of mice. *Proceeding of the National Academy of Sciences USA,* ***106***, 21335-21340.

[3] Spruijt, B.M., Peters, S.M., de Heer, R.C., Pothuizen, H.H.J., van der Harst, J.E. (2014). Reproducibility and relevance of future behavioral sciences should benefit from a cross fertilization of past recommendations and today's technology: "Back to the future". *Journal of Neuroscience Methods* (available online at http://dx.doi.org/10.1016/j.jneumeth.2014.03.001).

[4] Baarendse, P.J.J., van Grootheest, G., Jansen, R.F., Pieneman, A.W., Ögren, S.O., Verhage, M., Stiedl, O. (2008). Differential involvement of the dorsal hippocampus in passive avoidance in C57BL/6J and DBA/2J mice. *Hippocampus,* ***18***, 11-19.

[5] Stiedl, O., Radulovic, J., Lohmann, R., Palve, M., Birkenfeld, K., Kammermeier, J., Sananbenesi, F., Spiess, J. (1999). Strain and substrain differences in context- and tone-dependent fear conditioning. *Behavioural Brain Research,* ***104***, 1-12.

[6] Diamond, D.M., Campbell, A.M., Park, C.R., Halonen, J., Zoladz, P.R. (2007). The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity,* ***2007***, 60803.

[7] Stiedl, O., Birkenfeld, K., Palve, M., Spiess, J. (2000). Impairment of conditioned context- but not tone-dependent fear of C57BL/6J mice by the NMDA receptor antagonist APV. *Behavioural Brain Research*, ***116***, 157-168.

[8] Yang, M., Augustsson, H., Markham, C.M., Hubbard, D.T., Webster, D., Wall, P.M., Blanchard, R.J., Blanchard, D.C. (2004). The rat exposure test: a model of mouse defensive behaviors. *Physiology & Behavior,* ***81***, 465-473.

[9] Koolhaas, J.M., de Boer, S.F., Buwalda, B., van Reen, K. (2007). Individual variation in coping with stress: a multidimensional approach of ultimate and proximate mechanisms. *Brain, Behavior & Evolution,* ***70***, 218-226.

What's *always* wrong with my mouse?

N. Kafkafi, T. Lahav and Y. Benjamini

Department of Statistics and Operations Research, The Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
nkafkafi@gmail.com

In recent years there has been a growing voice of concern that a considerable percentage of published scientific discoveries fail to replicate in subsequent studies. The issue is especially relevant to preclinical studies and animal models, and has recently led to reconsideration of policies by NIH [1], as well as by some scientific journals including *Science* [2] and *Nature* [3]. Behavioral phenotyping results especially seem to be sensitive, and studies comparing inbred strains and genetically-engineered mutants across laboratories demonstrated some disturbing discrepancies [4]. These discrepancies are all the more worrying in light of the current community effort, coordinated by the International Mouse Phenotyping Consortium (IMPC), to phenotype thousands of mouse mutant lines across many laboratories during the next several years, and make the results available in public databases [5], as part of a long-term goal to functionally annotate all mammalian protein-coding genes.

While it is obvious that *something* should be done about the problem, it is less clear *what*. The new policies mostly advocate general methodological procedures and considerations, such as reporting detailed methods, pre-registering studies before the experiment and committing to sample sizes. However, they usually do not propose new statistical criteria and tools specifically designed to address the issue. Unfortunately, the intuitive notion of replicability as a central dogma of modern science has never been explicitly formulated. As recently noted by a statistician in a debate regarding replicating mouse phenotyping results: “The concept of reproducibility has not been well developed in the statistical literature, and so it is no wonder that debates like this have arisen” [6].

When estimating the difference between standardized mouse genotypes (e.g., an animal model knockout and its wild-type control) across several laboratories, the traditional criterion mostly used for a lack of phenotyping replicability is still the statistical significance of Genotype × Laboratory interaction ($G{\times}L$), although it is in fact misleading: it punishes high-quality behavioral measures in which the lower noise in measuring the individual animal effect (“within group”) increases the power to discover all other effects: Genotype, Laboratory and $G{\times}L$ [7]. The other side of this problem is that low-quality, noisy measures might mistakenly appear replicable, if they fail to discover $G{\times}L$ while just dis0063overing some strong genotype differences. We therefore argue that the more appropriate statistical model is mixed model ANOVA, in which the laboratories are regarded as a random variable (“random lab model” or RLM, as opposed to the traditional “fixed lab model” or FLM). RLM considers the laboratories in the study as *a sample*, representing the population of all potential phenotyping labs out there. It therefore adds the $G{\times}L$ “noise” to the individual animal noise as the yardstick against which genotype differences are judged. In practical terms, adopting RLM means raising the benchmark for showing a significant genotype effect, thus trading some statistical power for ensuring replicability [7].

In order to further examine the relevance of the FLM and RLM for replicability across laboratories we analyzed behavioral results from several mouse phenotyping studies, each conducted across several laboratories. Using these data we demonstrate that the commonlly-used FLM analysis frequently generates inconsistent conclusions that do not correspond with the intuitive concept of replicability. A typical example is seen in Figure 1, which shows a comparison between two genotypes, C57BL/6 and DBA/2, in the total path moved in the Elevated Zero Maze across 6 laboratories. FLM analysis indicated that C57BL/6 was significantly more active than DBA/2 across all laboratories (p<0.05), while RLM did not discover significant differences (p=0.47). Note that in 2 out of the 6 labs the DBA/2 mean was actually higher. Even worse, within one of these laboratories the DBA/2 mean was *significantly* higher, as indicated by the commonly-used t-test within this lab, although in 3 other labs it was significantly lower.

This kind of inconsistency in FLM analysis is not rare: in this dataset it was found in 30% of the measures in which genotype difference across all laboratories was significant in the FLM but not in the RLM. In contrast it was found in none of the measures in which RLM too indicated a significant effect across all laboratories. The same comparison in another dataset, the "heterogenized" dataset from Richter et al. [8], revealed an even worse result for the FLM: 40% vs. none.

In conclusion, our examination of the data reveals that the commonly-used statistical model, in which the laboratory is treated as a fixed variable, should not be used for estimating replicability of phenotyping results across laboratories. Instead we recommend using the significance of the genotype difference in a model that treats the laboratory as a random variable.

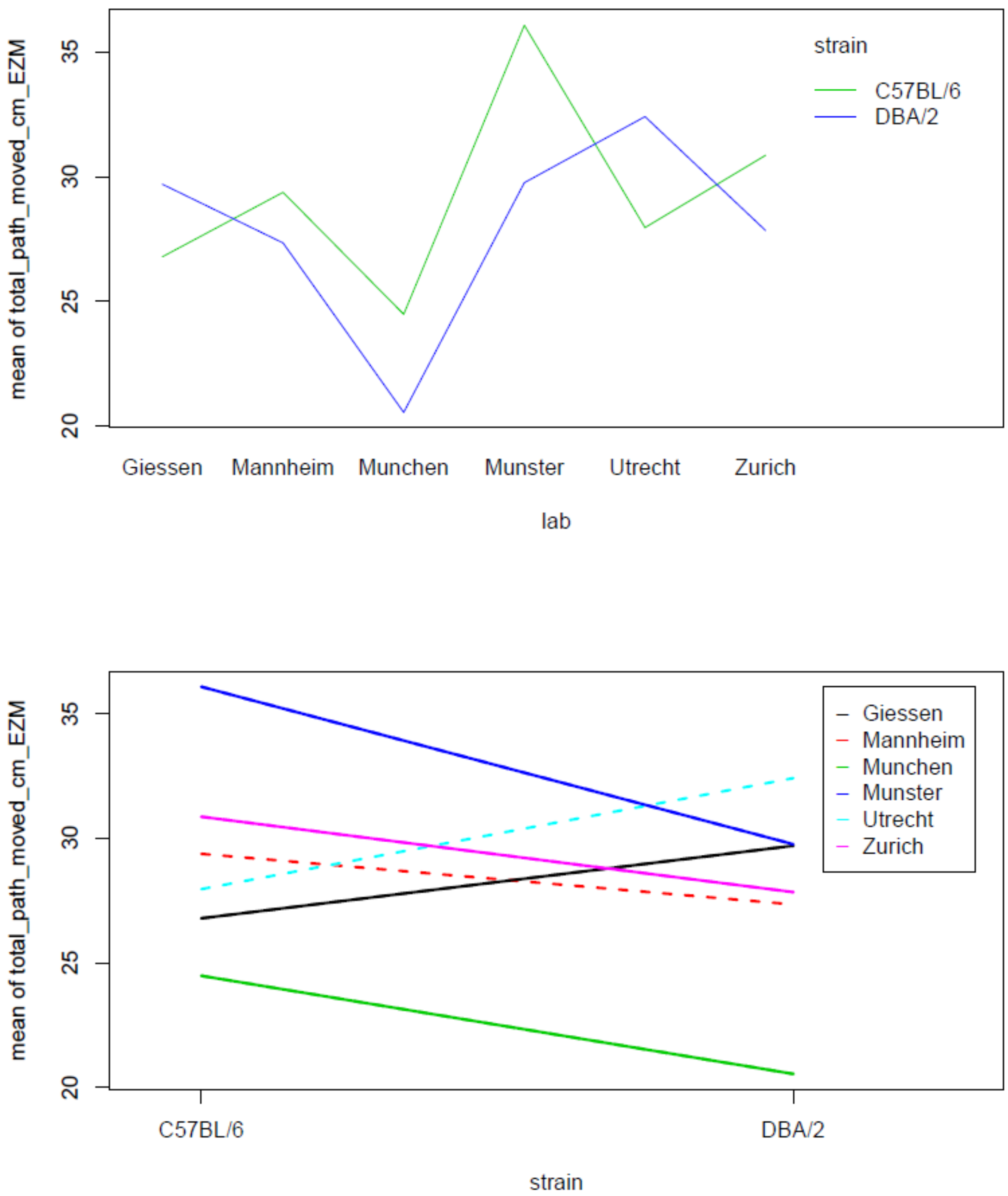


Figure 1. Differences between mouse genotypes C57BL/6 and DBA/2 across 6 laboratories, in the path moved in the Elevated Zero Maze, out of the Richter et al 2011 [8] "standardized" dataset. The same group means are are according to genotype (top) and to laboratory (bottom), and they are connected by lines in order to visualize Genotype × Laboratory interaction, seen as different and occationly even opposite slopes. In the bottom graph, continuous lines indicate genotype differences that were significant (p<0.05) using two-tailed t-test within the corresponding laboratory, and dashed lines denote non-significant differences.

Acknowledgements

This work is supported by a European Research Council grant PSARPS. The data from Richter et al. [8] and Wolfer et al. [9] were generously contributed to us by Prof. Würbel.

References

[1] Collins, F.S., Tabak, L.A. (2014). Policy: NIH plans to enhance reproducibility. *Nature, **505***, 612–613.

[2] McNutt, M. (2014). Reproducibility: *Science* Editorial. *Science, **343***, 229.

[3] Landis, S. C. et al. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. *Nature, **490***, 187–191.

[4] Crabbe J.C., Wahlsten D., Dudek B.C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science, **284***, 1670-1672.

[5] Koscielny, G. et al., (2014) The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Research.* **42**(D1): D802-D809.

[6] Wolfinger, E.D. (2013). Reanalysis of Richter et al. (2010) on reproducibility. *Nature Methods, **10***, 373–374.

[7] Kafkafi, N., Benjamini, Y., Sakov, A., Elmer, G.I., Golani, I. (2005). Genotype-environment interactions in mouse behavior: a way out of the problem. *Proceedings of the National Academy of Sciences USA, **102***, 4619-4624.

[8] Richter et al. (2011) Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PLoS One,* **6**, e16461.

[9] Wolfer, D.P., Litvin, O., Morf, S., Nitsch, R.M., Lipp, H.P., Würbel, H. (2004). Laboratory animal welfare: cage enrichment and mouse behaviour. *Nature, **432***, 821-822.

Estimating replicability of behavioral phenotyping results in a single laboratory

Y. Benjamini, T. Lahav and N. Kafkafi

Department of Statistics and Operations Research, The Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
nkafkafi@gmail.com

In studies that phenotype mouse inbred strains and mutants across several laboratories, the common way of estimating replicability is using the statistical model that treats both Genotype and Laboratory as fixed ("Fixed Lab Model" or FLM). As we demonstrated in previous work [1] this model is in fact misleading, and the more correct and intuitive estimation of replicability is given by a "mixed model" that treats Laboratory as a random variable ("Random Lab Model" or RLM). The RLM considers the laboratories in the study as a sample representing the population of all potential phenotyping laboratories out there. It therefore "inflates" the usual individual animal variability ("within-group") with the Genotype × Laboratory interaction ($G \times L$) variability, thus generating a higher benchmark against which the genotype differences are judged. In practical terms, adopting RLM means raising the benchmark for showing a significant genotype effect, thus trading some statistical power for ensuring replicability.

It is important to realize, however, that in practice almost all phenotyping studies are currently done in single-lab studies. Suppose a researcher phenotypes, in her own lab, a mutant line considered to be an important animal model, and makes "a discovery" – the difference between the mutants and their wild-type controls is statistically significant. How would she know if this significant effect is likely to replicate in other labs? Should she publish the discovery, or seek first to validate it in additional labs? How would other researchers know whether to use this mutant in their lab, expecting to observe similar results?

Kafkafi and Benjamini et al. [1] first proposed using the RLM in multi-lab studies to estimate the size of $G \times L$ variance for each standard phenotyping measure, and then use it to "inflate" the variance of any single-lab experiment employing the same measure, even with different genotypes, to enable the researcher to test whether her discovery is likely to replicate in other labs. We examined this proposition by analyzing behavioral results from several mouse phenotyping studies, each conducted across several laboratories over the world. The data from Richter et al. [1] and Wolfer et al. [3] were generously contributed to us by Prof. Würbel.

For each phenotyping measure we first established, using RLM, whether a significant genotype effect is found across all laboratories in the study, and used it as our indicator for true replicable genotype effect. We can thus examine if a researcher in one of these labs can use only her local results to discover these effects, while avoiding false discoveries. We compare the standard way of doing so – two-tailed t-test using the standard error measured in lab – with our proposed method of "inflating" the locally estimated standard error by the estimated $G \times L$ standard deviation. In the intended use of this method, $G \times L$ will be estimated from previous multi-lab experiments using the same measures, but here we do not have such experiments, so we use instead the $G \times L$ standard deviation as can be estimated from the analysis over all laboratories.

Figure 1 (top) shows a case of very significant genotype effect according to both FLM and RLM. The similar slopes of the 6 lab lines indicate small $G \times L$. The different vertical heights of the lines indicate a laboratory effect, but this is not a serious issue since by definition it is common to all genotypes in each lab, and therefore does not affect the difference between them. Unfortunately, such clear-cut cases are rare: out of 29 measures in this dataset, only 6 had non-significant $G \times L$. As noted by Crabbe et al. [4], significant $G \times L$ is the more common case.

Cases in which the data from all labs do not indicate a true genotype difference enable us to quantify false discoveries. Figure 1 (bottom) is a striking example of the failure of the traditional method, which claimed here significant genotype differences in 4 out of the 6 labs. Moreover, in one of these labs the significant effect was actually in the opposite direction to the other 3. In contrast, our proposed method based on the RLM did not detect a significant effect in any of these laboratories. In this particular case the traditional method therefore suffered a false discovery proportion of 4/6 = 66.7%, compared to 0% in our proposed method.

Measures in which the multi-lab analysis indicates a true genotype effect, as in Figure 1 (top), enables us to compare the statistical power of the two methods. The traditional method discovered the true genotype difference in all 6 labs, thus scoring here a statistical power of 100%, while our proposed method missed the difference in one lab, in which the effect was not strong enough to overcome the slightly inflated yardstick for significance, and therefore scored a statistical power of just 5/6 = 83.3%. As expected, our more conservative proposed method avoids many false discoveries but pays for that with a somewhat reduced power to discover real differences. Over this entire dataset, our proposed method decreased the false discovery proportion from 32.3% to 13.3%, compared to the traditional method, at the price of reducing the power from 74.4% to 43.3%.

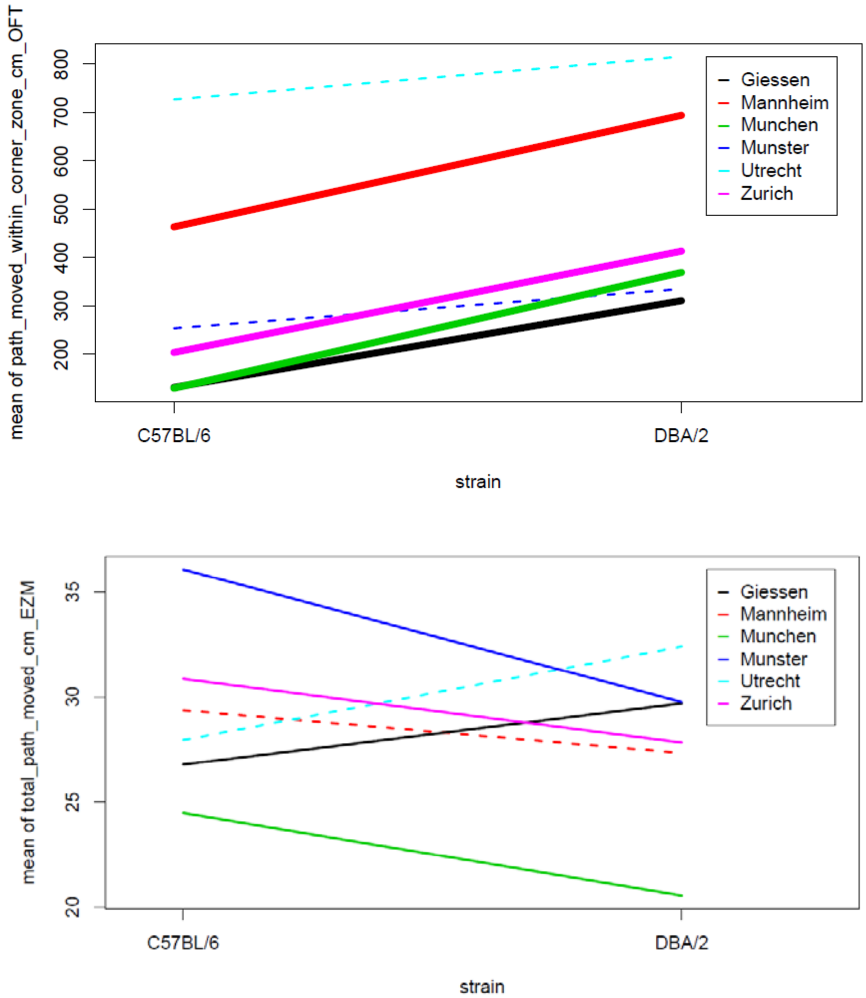


Figure 1. Differences between genotypes C57BL/6 and DBA/2 across 6 laboratories, in three measures out of the Richter et al 2011 “standardized” dataset: path moved within corners in the Open Field Test (top), Time spent in exploration zone in the Novel Object Test (middle), and path moved in the Elevated Zero Maze (bottom). Each line connects genotype means within the laboratory. Dashed lines: non-significant within the laboratory; continuous thin lines: significant within the lab according to two-tailed t-test; bold lines: significant within the lab after “inflating” the in-lab results with the G×L variability measured across all laboratories.

In conclusion, the proposed method is recommended for estimating if phenotyping results within a single laboratory are likely to replicate in other laboratories, assuming a previous estimation of the interaction was already established from several laboratories. Public phenotyping databases, e.g., of the International Mouse Phenotyoing Consortium [5], can store $G \times L$ variability for each phenotyping measure, so that any researcher can easily use it to determine the replicability of discoveries in her single-lab study. The commonly-used criterion for declaring phenotyping discoveries in a single-lab study – the significance of the genotype effect using within-lab variability only – probably generates a high percentage of "discoveries" that will not replicate in other laboratories.

Acknowledgements

This work is supported by a European Research Council grant PSARPS. Data was kindly contributed by Prof. Hanno Würbel.

References

[1] Kafkafi, N., Benjamini, Y., Sakov, A., Elmer, G.I., Golani, I. (2005). Genotype-environment interactions in mouse behavior: a way out of the problem. *Proceedings of the National Academy of Sciences USA*, ***102***, 4619-4624.

[2] Richter et al. (2011). Effect of population heterogenization on the reproducibility of mouse behavior: A multi-laboratory study. *PLoS One*, ***6***, e16461.

[3] Wolfer, D.P., Litvin, O., Morf, S., Nitsch, R.M., Lipp, H.P., Würbel, H. (2004). Laboratory animal welfare: cage enrichment and mouse behaviour. *Nature*, ***432***, 821-822.

[4] Crabbe J.C., Wahlsten D., Dudek B.C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science,* ***284***, 1670-1672.

[5] Koscielny, G. et al., (2014) The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Research*, ***42(D1)***, D802-D809.

The demand for replicability of behavioral result: from burden to asset

Ilan Golani[1], Yair Wexler[2] and Yoav Benjamini[2]

[1]Department of Zoology and Sagol School for Neuroscience, Tel Aviv University, Tel Aviv, Israel
ilan99@post.tau.ac.il
[2]Department of Statistics and Operatio Research and Sagol School for Neuroscience, Tel Aviv University, Tel Aviv, Israel

Introduction

The demand for replicability of behavioral results across laboratories is viewed as a burden in behavior genetics. We demonstrate how, by using replicability as a design concept, we turn it into an asset, offering a quantitative criterion that guides the design of better ways to describe behavior. In this presentation we focus on how we use replicability for the design of better quality building blocks of exploratory behavior. Passing the high benchmark dictated by the replicability demand and at the same time obtaining measures that have sufficient discriminative power requires higher quality data. Some of the procedures we use to obtain such data include measuring behavior in reference to natural frames of reference and natural origins of axes used by the animals themselves, using individually customized cutoff points between the building blocks of behavior, and selecting those building blocks that fulfill the criteria for replicability, all the while filtering out building blocks that are not replicable across laboratories.

Sorting incursions into replicable and non-replicable types

We will demonstrate this procedure with a single example. Direct observation of mouse open field behavior suggests that arena wall is used by the mice as a natural origin of axes for the performance of incursions - forays into arena centre that start and end at the wall. Given this observation it only makes sense to use the wall as the origin of axes for the measurement of incursions. Observation further suggests that the depth of incursion into the center is a relevant parameter of this behavior. Therefore, it only makes sense to classify incursions according to the maximal distance they reach from the wall. The empirical density distribution of maximal distance from wall of incursions (see Figure 1a) suggests that the population of incursions consists of a mixture of three sub-populations. This impression is supported by fitting a Gaussian mixture model to the empirical density function (see Figure 1, from [1]).

Without a demand for replicability and a statistical yardstick for measuring it our study would end up by adding three incursion types to the mouse's repertoire ("ethogram"). A test for replicability endows us with the absolutely necessary function of examining the validity of these three novel building blocks beyond the reality of our own lab and the function of revealing differences across strains, treatments and preparations. With regard to incursions, testing for replicability reveals that only two of the incursion types are replicable across laboratories in the forced open field test (see Figure 2).

As shown, a comparison of the numbers of Incursions performed during a session in, for example, C57BL/6J and DBA/2J mice in 3 labs, shows that the number is higher in the first strain, but this difference is not statistically significant (see Figure 2, top left panel). Scoring of the three incursion types that are isolated by classifying incursions according to their maximal distance from wall (see Figure 1) is plotted in Figure 2 (top right) and two bottom panels. The numbers of near-wall incursions are evidently not replicable across laboratories: although there is a large strain difference in TAU, in the other two laboratories the two strains have a similar number of near-wall incursions. In contrast, there is a replicable strain difference in the number of intermediate and arena-crossing incursions, with C57BL/6J mice making significantly more incursions of both these types than DBA/2J mice in all three laboratories. It now becomes evident that the failure to achieve significant results in the overall number of incursions (See Figure 2, top left) is due to inter-laboratory variation in the numbers of a single incursion type: near-wall incursions. The boxplot summaries disclose how stratified scoring plus a multi-laboratory experiment transform lesser quality measures into higher-quality measures: we start with non-significant differences in the un-stratified measure (See Figure 2, all incursions; partly overlapping spread of the boxplots of the two strains in all three laboratories), and continue with filtering out the segment type that shows a strong interaction (near-wall incursions), and we end up by keeping two new replicable measures, that of Number of Incursions in intermediate and that of Number of Incursions in arena-crossing incursions (that show a similar, consistent difference between the two strains' boxplots in all three laboratories).

Summary

the current outcry with regard to the lack of replicability of behavioral phenotyping results highlights only one aspect of the crisis, having to do with the poor predictive value of scientific results that reflect inappropriate handling of data [2][3]. The other aspect, also contributing to poor predictive value of results, is largely ignored. It has to do with the poor quality of the measures that are used to establish phenotypic differences. By judiciously selecting candidate measures (e.g., aspects of behavior that are suspected to be performed in reference to candidate reference values selected by the organism itself, as, for example. incursions that seem to be performed by mice in reference to arena wall), and by judicious preparation of the data for analysis (e.g., smoothing [4] and segmentation [5] based on intrinsic statistical and geometrical features of the data) we increase the likelihood of obtaining predictive results. It is, however, a proper test for replicability that validates or refutes our initial selections, improving the universal status of measures and building blocks that receive a high replicability score [6].

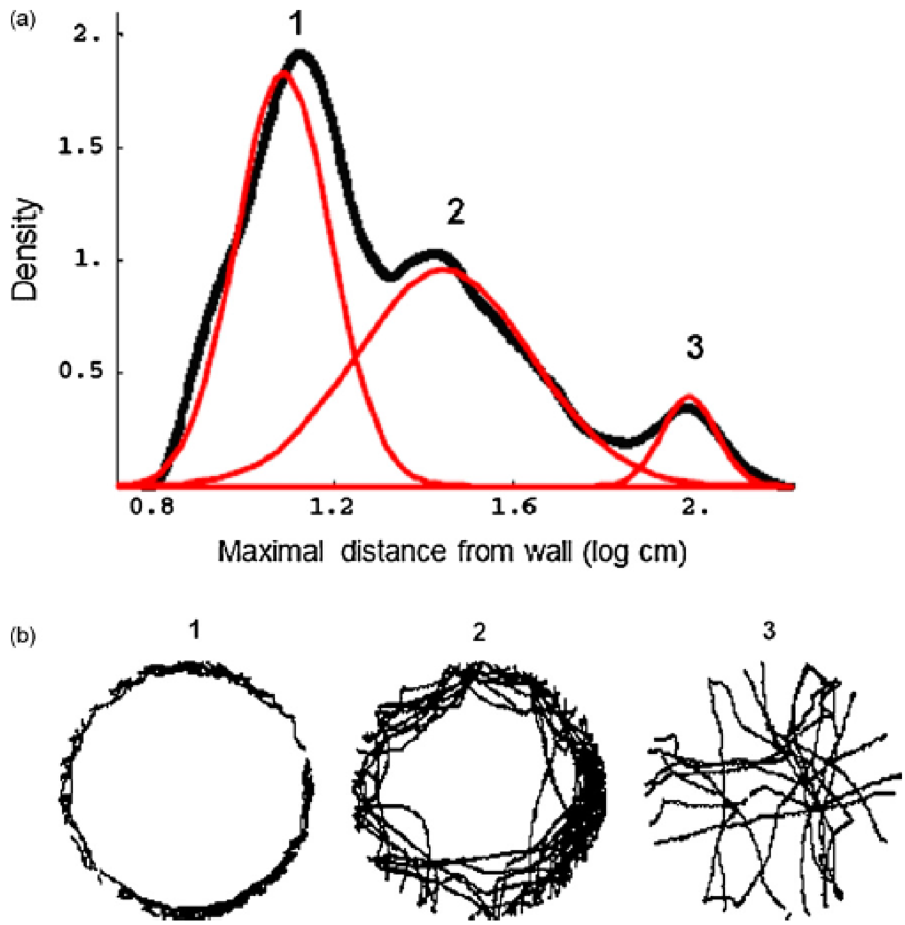


Figure 1. Black: a density graph of the distribution of the maximal distances from wall of centre segments (log transformed) in a single C57BL/6J session. Red: three Gaussians fitted to the distribution by the EM algorithm. The intersection points between the Gaussians serve as cutoff values for dividing all incursions performed in this session into three types. (b) Path plots of the incursions belonging to each type (from[1]).

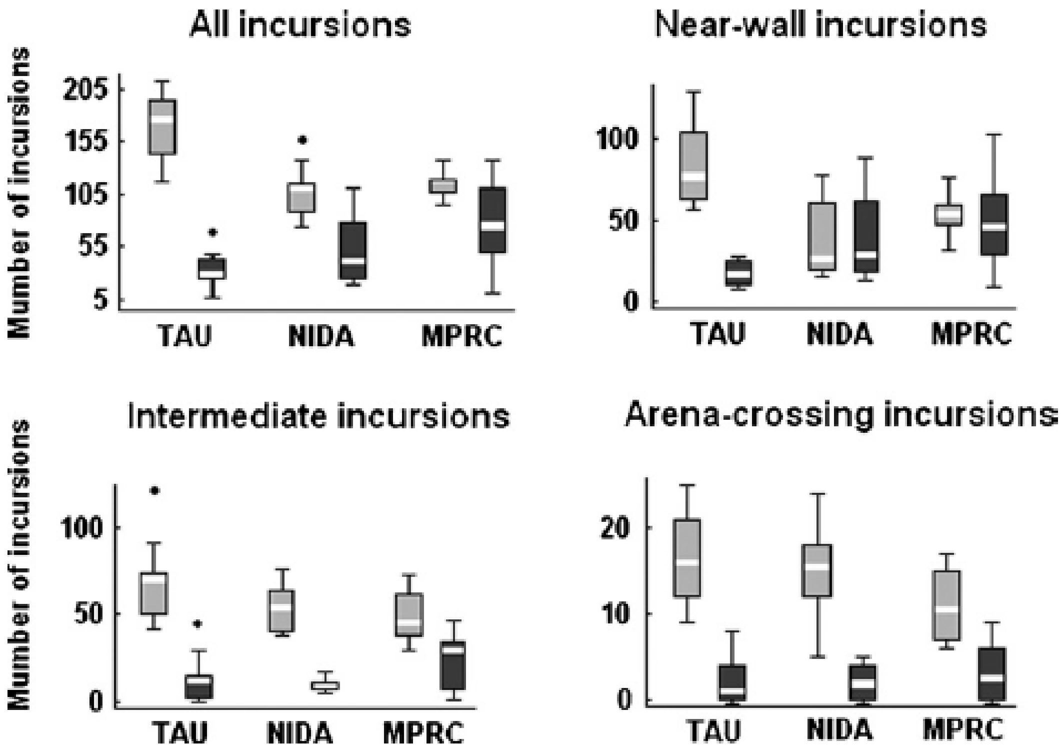


Figure 2. Comparison of the degree of replicability of the number of incursions performed during a session in DBA/2J (in gray) and C57BL/6J mice (in black) before and after stratified measurement. Top left: measurement of all incursions pooled together (before stratification). The differences between the strains are not consistent (nonreplicable) across laboratories. Top right and two bottom panels: After stratified measurement, the difference between the strains across laboratories becomes more consistent in intermediate and in arena-crossing incursions (this is represented visually in the consistent vertical distances between boxplots positions in the graphs). It is the non-replicable difference between near-wall incursions in the 2 strains that masked the highly replicable differences in the other 2 incursion types (from [1]).

Acknowledgements

This study was funded by a European Research Council grant PSARPS to YB.

References

[1] Lipkind, D., Sakov, A., Kafkafi, N., Elmer, G.I., Benjamini, Y., Golani, I. (2004). New replicable anxiety-related measures of wall vs center behavior of mice in the open field. *Journal of applied physiology (Bethesda, Md. : 1985)*, ***97****(1)*, 347–359.

[2] Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS medicine*, ***2****(8)*, e124.

[3] Cumming, G. (2008). Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, ***3****(4)*, 286–300.

[4] Hen, I., Sakov, A., Kafkafi, N., Golani, I., Benjamini, Y. (2004). The dynamics of spatial behavior: how can robust smoothing techniques help? *Journal of Neuroscience Methods*, ***133****(1–2)*, 161–172.

[5] Drai, D., Benjamini, Y., Golani, I. (2000). Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods*, ***96****(2)*, 119–131.

[6] Kafkafi, N., Benjamini, Y., Sakov, A., Elmer, G.I., Golani, I. (2005). Genotype–environment interactions in mouse behavior: A way out of the problem. *Proceedings of the National Academy of Sciences of the United States of America*, ***102****(12)*, 4619–4624.

Mouse Phenome Database: A resource to address phenotypic heterogeneity, research reproducibility and replicability

M.A. Bogue, S. C. Grubb, V. Philip and E.J. Chesler

The Jackson Laboratory, Bar Harbor, Maine, USA
molly.bogue@jax.org

Mouse Phenome Database (MPD; phenome.jax.org) allows data from complex trait studies to be found, reproduced, reused and placed into the context of other studies

Access to primary data is crucial for three reasons: 1) reproducibility, 2) reanalysis in light of new developments and 3) integrative analysis to find consensus among diverse studies. Unfortunately, data often exist in diverse and sometimes non-computable stores with insufficient documentation and restricted access. MPD is a widely used online resource providing access to primary experimental data and protocols in the predominant genetic model organism, the laboratory mouse [1]. The MPD, in existence for the past 12 years, amasses, annotates, integrates and maintains primary quantitative phenotype data and protocols in a centralized public database. Since the inception of MPD, a wealth of phenotype technologies and mouse resources have led to an expanded scope and refocus of the system – from inbred strain characteristics to a rigorously curated data resource for complex trait and integrative genetic analysis. This resource houses phenotypic, gene expression or genotype data for >1300 strains. MPD provides a catalog of phenotypic assays, analysis tools to explore genetic variation, and a common framework for data access and data dissemination. Data come from investigators around the world (supported by ~130 governmental funding agencies and research foundations) and represent a broad scope of behavioral endpoints and disease-related characteristics in naïve mice and those exposed to drugs, environmental agents or other treatments. MPD provides an important venue for compliance with data sharing policies and facilitates data reuse, saving time and resources while reducing animal use.

Integrating various sources of phenotype data in MPD provides researchers with the resources they need to reproduce experiments, reanalyze genetic studies with new algorithms and genetics maps, understand relationships among traits and elucidate the shared genetics for a multitude of traits. The high level of documentation and curation standards and stability of the program at The Jackson Laboratory have made MPD a primary resource for investigators to archive and retrieve quantitative mouse phenotypic data. This high-quality, standardized data resource enables investigators to select mouse strains for modeling disease, compare results of diverse phenotypic assays and benchmark experimental data using detailed protocols.

The utility of model organism research is dependent on excellent research reproducibility

Inadequate reporting and encumbered access to primary data reduce the impact of studies and act as a barrier to effective translation of scientific discoveries from basic/preclinical findings to human applications including development of new diagnostics and therapies. Substantial recent attention has been given to poor cross-species relevance, replicability and reproducibility of research findings in the laboratory mouse. The reasons for poor reproducibility are many-fold and include inadequate reporting of mouse strains and resources, insufficient statistical power, diversity among experimental protocols and lack of documentation of research resources. It is imperative that studies are reported with detailed experimental information to allow researchers to evaluate reported results, repeat experiments and extend findings.

MPD has consistently provided rigorous curation of mouse experimental data to help alleviate issues associated with reproducibility. By structuring mouse phenotyping studies, annotating them to controlled vocabularies and developing integrative tools that rely on the unique value of this data, MPD facilitates access and reuse of primary phenotype data, enabling cross-species comparisons and ultimately assuring relevance to human studies.

This will help achieve a more holistic understanding of disease processes and mechanisms and maximize the value of the data while leveraging the investments made in basic research.

Interpreting the heterogeneity of behavior

Understanding the genetics of complex disease, and in particular behavioral disorders, requires sensitivity to the challenge of modeling disease in mice in a manner that captures the heterogeneity, complex interrelations and co-occurrences among disorders. Behavioral experiments are quite sensitive to environmental conditions and experimenter effects. Furthermore, researchers often choose a single experimental assay as an endpoint among many possible tests of a behavioral characteristic, despite the different biases or varied aspects of behavior modeled by each. Inconsistencies among these assays may confound the already deep complexity and heterogeneity of behavior. Although mice cannot recapitulate the full complexity of certain psychiatric disorders and diseases, many component features can be independently observed and genetically characterized.

MPD tools help investigators to identify a consensus among diverse assays and varied environments through genetic correlations among traits and underlying molecular mechanisms. This strategy is widely used in behavioral genetics and has made behavior one of the most highly represented fields of study in MPD. We are in the process of enhancing the MPD system by expanding the scope of data and developing new analysis tools to further address phenotypic heterogeneity, research reproducibility and replicability. We have plans to incorporate a more sophisticated suite of tools for integrated multi-dimensional analysis that will better enable correlation analyses and discovery of biological mechanisms. Genetic parameter estimations and population distributions will help researchers to identify those assays and mouse resources that will provide the best experimental characteristics for the assessment of specific behavioral and biological concepts. Integrative analysis methods will enable discovery of coherent signal amidst gene-environment interactions, laboratory environmental diversity and assay diversity.

Reference

[1] Grubb, S.C., Bult, C.J., Bogue, M.A. (2014). Mouse Phenome Database. *Nucleic Acids Research, **42**(Database issue)*, D825-834.

Automatized tracking of free-ranging wild song birds using the Encounternet

Marc Naguib and Lysanne Snijders

Behavioural Ecology Group, Wageningen University, Wageningen, The Netherlands

Understanding the ecology and evolution of animals requires fundamental insights into the decision making of individuals. Animal populations are characterized by individuals usually moving in space and time to explore and exploit resources, find partners, or escape predators. Thus many decisions they make have a spatial component, potentially affecting reproduction and survival. Decisions where to move when has implications for access to information about the environment and for encountering predators, potential mates and competitors and thus for the social dynamics. Yet, despite a profound understanding on selection pressures on animals, we still have very little information about individual spatial activities as it is very difficult to track free ranging animals over a prolonged period of time. Even though GPS devices can provide high accuracy tracking data, they are limited to larger animals. Also novel light-weight GPS tags are limited for continuous tracking, due to their (still) low battery life time (Wikelski *et al.* 2007). For small animals, radio-tracking has been a successful alternative as light weight senders can last for several weeks. Yet, radio tracking usually requires a person to follow a given individual and mark its locations at regular time intervals which is a time consuming expedition specifically for mobile small animals, such as songbirds (Naef-Daenzer 1994; Naguib *et al.* 2001; Amrhein *et al.* 2004; Roth *et al.* 2009).

Songbirds are a key model organism for research on the ecology and evolution of vertebrate behaviour. Studies on the song of the males as well as less variable traits, like colour ornaments have provided important insights into how selection acts on these traits and in many species reproductive success can be readily determined, allowing to link behaviour and life history traits to environmental factors and fitness. Yet, due to the difficulty of following individuals over prolonged periods of time, many of the behavioural mechanisms underlying fitness effects are not well understood. Where and when do individuals forage? How do they explore their environment and how are the spatial associations among neighbouring birds? Are females near males when these males sing, and if so, how long do they stay nearby and listen? Advances in understanding such questions require to follow individuals, at best, continuously in space and time, allowing to also determine individual's position within a social network (Croft *et al.* 2007).

By using the Encounternet (Burtsoft, Portland, Oregon), a state-of-the-art technology of digital radio tags (Mennill *et al.* 2012), combined with a grid of receivers placed at the study site we were able to collect unprecedented data on spatial movements of several individuals at the same time and their social connectivity in a natural population of wild great tits (*Parus major*). Great tits are resident, non-migratory songbirds. During their breeding season they are territorial, yet have large home ranges, encountering each other regularly (Hinde 1952). These novel tracking techniques allow to determine home ranges of many individuals simultaneously and to determine which individuals meet where and when.

In this presentation we will provide insights into the functioning of this tracking system and show data on spatial movements and social network structure of great tits.

References

[1] Amrhein, V., Kunc, H.P., Naguib, M. (2004). Non-territorial nightingales prospect territories during the dawn chorus. *Proceedings Royal Society B*, ***271***, S167-S169.

[2] Croft, D.P., James, R., Krause, J. (2007). *Exploring animal social networks*. Princeton: Princeton University Press.

[3] Hinde, R.A. (1952). The behaviour of the great tit (*Parus major*) and some other related species. *Behavioural Processes*, ***2(Supplement)***, 1-201.

[4] Mennill, D.J., Doucet, S.M., Ward, K.A.A., Maynard, D.F., Otis, B., Burt, J.M. (2012). A novel digital telemetry system for tracking wild animals: A field test for studying mate choice in a lekking tropical bird. *Methods in Ecology and Evolution*, ***3***, 663-672.

[5] Naef-Daenzer, B. (1994). Radiotracking of great and blue tits: new tools to assess territoriality, home-range and ressource distribution. *Ardea*, ***82***, 335-347.

[6] Naguib, M., Altenkamp, R., Grießmann, B. (2001). Nightingales in space: song and extra-territorial forays of radio tagged song birds. *Journal of Ornithology*, ***142***, 306-312.

[7] Roth, T., Sprau, P., Schmidt, R., Naguib, M., Amrhein, V. (2009). Sex-specific timing of mate searching and territory prospecting in the nightingale: nocturnal life of females. *Proceedings Royal Society London B*, ***276***, 2045-2050.

[8] Wikelski, M., Kays, R.W., Kasdin, N.J., Thorup, K., Smith, J.A., Swenson Jr., G.W. (2007). Going wild: what a global small-animal tracking system could do for experimental biologists. *Journal of Experimental Biology*, ***210***, 181-186.

Miniature wireless microphones reveal vocal relationships in a group of small laboratory animals

A. Ter Maat*, L.Trost, H. Sagunsky, M. Gahr

Max Planck Institute for Ornithology, Seewiesen, Germany

* termaat@orn.mpg.de

We have developed lightweight wireless microphones that can be carried by small animals. The microphones weigh 0.6 g including battery, run for two weeks and have a range of over 5 m, sufficient to be used in an indoor aviary (Figure 1). We operate the device without an external antenna in order to minimize interference with behavior.

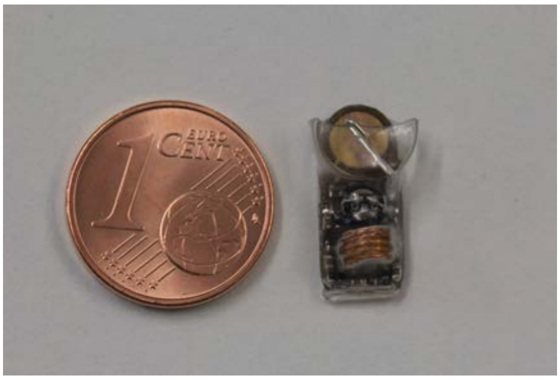


Figure 1. Wireless microphone used to record individual vocalizations.

The spread in transmitting frequency is used to enable the simultaneous recording of up to 12 animals. For each animal one needs a separate receiver. Transmitting frequency is about 300 MHz. All experiments were done indoors. Crossed yagi antennae (Winkler Spezialantennen, Kreuzdipol 300, directional antenna tuned at 300 MHz) were connected to AOR 8600 receivers (AOR, Ltd., Japan) that were modified to an audio bandwidth of 12 KHz. Signals are received as FM, at an intermediate frequency bandwidth of 100 KHz. The antennae were less sensitive to changes in orientation of the transmitters relative to the antennae, for instance when the animal moved. To compensate for frequency drift due to a change in battery voltage or temperature fluctuations, the frequencies of the transmitters were tracked by the receivers using a custom-built software based tracking system.

The device is used to obtain recordings from each individual in a group. By mounting the device such that the (actual) microphone component faces towards the animals' bodies, a recording is obtained that selectively records the vocalizations of each bird (Figure 2).

Vocal interactions were analyzed as follows. The output of the receivers was stored as wave files. The vocalizations were identified based on acoustical parameters using custom software written in Delphi Pascal and C++ ((Jansen et al., 2005). Different types of vocalization were sorted and assigned cluster numbers that were time-stamped. Based on this information the temporal relationships between the sounds could be established using peristimulus time histograms that essentially measure the probability of one call occurring before or after the call of another animal (Figure 3;[1]).

Figure 2. Selectivity of the wireless microphone. Upper trace: recording of 6 animals in an aviary using a standard microphone. The wireless microphone (lower trace) records the vocalizations of one animal only. This allows the correlation of individual vocalization patterns to describe the vocal interactions within this group.

As an example of the uses to which the wireless microphones can be put we report on the vocal communication of a highly social songbird, the zebra finch. In the zebra finch, song is essentially a one-way vocal communication pathway that is shaped by sexual selection [2]. Zebra finches also produce large numbers of soft, unlearned calls, among which "stack" calls are uttered frequently [3]. The social function of these calls is, however, poorly understood. To chart the vocal interactions between the individuals in a group, it is critical to be able to determine unequivocally the calls produced by every animal. We achieved this by mounting miniature wireless microphones on all individuals (Figure 2). This revealed that group living males and females communicate using bilateral stack calling. This vocal pattern occurred predominantly between bonded partners.

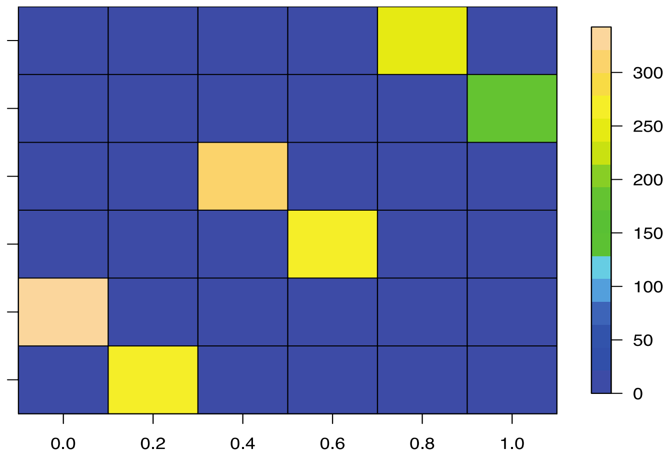


Figure 3. Mapping out social interactions based on peristimulus time histograms. Here, the strength of the interaction, disregarding direction, is shown for three bonded pairs in an aviary.

Using quantified measures of direction and strength of the vocal interactions a model of the social structure can be developed. We quantified the intensity of the vocal interactions in groups of 2 to 4 pairs kept in an aviary (Figure 3). Response strength calculation is based on a PSTH consisting of 2 x 80 bins of 50 msec. General response strength:

$$RS = \frac{(N_{before} + N_{after}) - (N_{basebefore} + N_{baseafter})}{(N_{before} + N_{after}) + (N_{basebefore} + N_{baseafter})}$$

where N_{before} and N_{after} are the counts in the 9 bins before and after the start of the source event (=call) and $N_{basebefore}$ and $N_{baseafter}$ are the first and last 9 bins in the PSTH.

The strength and direction of the interaction were calculated on the basis of the cross-correlations between the calls of the individual birds. In this example we saw interactions only among bonded partners, which is the predominant pattern.

To our knowledge, we are the first to conduct a detailed study of the intricate vocal interactions in social groups of small songbirds.

Ethics statement

Carrying the transmitter as a backpack and keeping the birds in aviaries was approved by the government of Upper Bavaria, “Sachgebiet 54 – Verbraucherschutz, Veterinärwesen, 80538 München” with the record number: Az. 55.2-1-54-231-25-09. All further animal husbandry or handling was conducted according to the directives 2010/63/EU of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes.

References

[1] Abeles M (1982) Quantification, smoothing, and confidence limits for single-units' histograms. *Journal of Neuroscience Methods*, ***5***, 317-325.

[2] Gahr, M. (2007). Sexual differentiation of the vocal control system of birds. *Genetics of sexual differentiation and sexually dimorphic behaviors*, ***59***, 67-105.

[3] Zann, R.A. (1996). *The Zebra Finch*. Oxford: Oxford University Press.

Long-term behavioural repeatability in wild Eurasian perch (*Perca fluviatilis*) in a natural lake: A high-resolution biotelemetry study

S. Nakayama[1,2,*], K. Laskowski[1], P. Doering-Arjes[2], T. Klefoth[1] and R. Arlinghaus[1,2]

[1]Department of Biology and Ecology of Fishes, Leibniz-Institute of Freshwater and Inland Fisheries, Berlin, Germany.
[2]Faculty of Life Sciences, Humboldt-Universität zu Berlin, Berlin, Germany.
[*] shinn407@gmail.com

Consistent behavioural differences among individuals over time, also known as 'personality', have been discovered in many animals in a wide realm of behaviours [1]. To identify whether individual animals exhibit consistent individual differences in behaviour, repeated measures of the same behaviour are needed. Most personality studies achieve this through repeatedly measuring behaviour in standardized laboratory conditions over fairly short time intervals (one to a few days). However, these measures should be interpreted with caution if repeatability fluctuates over time, or varies with the number of observations. Moreover, behaviour can be especially sensitive to the external environment and the field is still lacking high-resolution repeated measurements of individual behaviour in natural settings. In our study, we used a unique long-term data set on fish behaviour in the field to investigate (1) whether behavioural repeatability is stable over time, (2) whether repeatability changes with temporal intervals at which the behaviour was measured, and (3) whether repeatability changes with the number of observations.

To accomplish these objectives, we tracked wild Eurasian perch (*Perca fluviatilis*) in a natural lake over 12 months using a high-resolution 3D acoustic telemetry system. The Kleiner Döllnsee (~25 ha) is a dimictic, shallow (mean depth 4 m, maximum depth 8 m), and slightly eutrophic lake, located 80 km northeast of Berlin, Germany. We installed 20 acoustic hydrophones in the lake, and recorded the positions of 16 adult female perch at high rates (up to 9-s intervals) for one year (from September 2009 to September 2010). The experiments were approved through an animal care permit (23-2347-15-2010) granted by the Ministry of Environment, Health and Consumer Protection Brandenburg, according to the German Animal Protection Act.

For each individual, we calculated daily mean values of swimming speed (m/s), turning rate (radian/s), and depth of fish (m). For each behavioural variable, we estimated repeatability as $r = s_A^2/(s^2 + s_A^2)$, where s_A^2 is the variance among individuals and s^2 is the variance within individuals, using a linear mixed model with individual as a random effect (MCMC method). We found that repeatability for a whole study period was comparable to previously published data [1] (swimming speed 0.176, turning speed 0.474, depth 0.160). However, behavioural repeatability differed between seasons, with higher repeatability found in winter compared to summer. To further investigate the temporal change of repeatability, we estimated repeatability every six consecutive days over a year, sliding by one day. Repeatability fluctuated markedly during the study period (for example, 0.262-0.822 for swimming speed). We then performed a linear model (with an autocorrelation term) on the repeatability to see whether abiotic environmental variables explain the temporal trend in repeatability. The model parameters were averaged over the models that constituted up to 95% model probability, and the significance of parameters were estimated using hierarchical partitioning. As a result, the fluctuation of repeatability was well explained by a number of environmental factors. For example, fish showed higher repeatability in swimming speed when water temperature was lower and light intensity was higher, but the repeatability was not explained by temperature fluctuation, air pressure, or precipitation. Secondly, to investigate the effect of temporal intervals on repeatability, the repeatability was calculated between randomly sampled two days. Repeatability declined with the interval between days, but the effect was very small. Finally, to investigate the effect of the number of observations on repeatability, we resampled the data and recalculated repeatability with different numbers of observations. We found that repeatability was underestimated when the observation number was small, but stabilized with increasing observations.

Our study is unique in offering an extremely high-resolution measurement of individual behaviour in the wild providing detailed insight into the temporal patterns and the influence of sampling on behaviour. These results demonstrate that personality of animals is expressed in nature and is rather stable over time, but that the description of this behavioural variation can be influenced by the time interval between and number of observations.

Reference

[1] Bell, A.M., Hankison, S.J., Laskowski, K.L. (2009). The repeatability of behaviour: A meta-analysis. *Animal Behaviour*, ***77**(4)*, 771-783.

Validation of the TrackLab Positioning System in a Cow Barn Environment

L. Frondelius[1], M. Pastell[1] and J. Mononen[1,2]

[1]Animal Production Research, MTT Agrifood Research Finland, Maaninka, Finland.
firstname.lastname@mtt.fi
[2]Department of Biology, University of Eastern Finland, Kuopio, Finland.
firstname.lastname@uef.fi

Introduction

Position data provides the means to track cows' activity and movement in a barn. This information may, for example, help in heat detection [1] or allow social relations in a herd to be studied [2]. We can identify the following requirements for a cow positioning system: The accuracy should be less than 1 m [3] and the sampling rate should be high enough [3], for measuring e.g. cow's speed. The system should be able to track multiple cows simultaneously [2].

However barn environment also has certain challenges::

- There are lots of metal structures in barn environment that may cause reflections of signal and artefacts [2, 3].
- Hardware, especially tag attached to the cow, should be dust- and water-resistant [2].
- Tag should be ergonomic and small [2]. It should not prevent animal's normal movement or behaviour.
- Long battery life [2]. However, often battery life in positioning tags is short (e.g. 24 h in [1]) and higher sampling rate means higher energy consumption and decreased battery life [2, 3].

TrackLab (Noldus Information Technology, Netherlands) is a commercially available system for recognition and analysis of spatial behaviour. It collects location data with Ubisense real time location system (Ubisense, Germany) which is based on Ultra Wide Band (UWB) radio signals. Animals are monitored by means of a tag (Ubisense Series 7000 Industrial tag) that transmits UWB pulses of extremely short duration and remote sensors (Ubisense Series 7000 IP Sensors) which enable location to be mapped by using Time-Difference-of-Arrival (TDoA) and Angle-of-Arrival (AoA) techniques. Maximum sampling rate is 137 Hz. MTT CowLab® is the first barn environment where the TrackLab system has been installed and tested. With six sensors it enables tracking the position of 50 cows simultaneously with accuracy of 30 cm.

The aim of this study was to validate the accuracy of the Tracklab system in a real barn environment. Here we present the result from the first TrackLab validation experiment in our cow barn.

Materials and methods

The positioning system was validated with a single measuring point test (SMPT) for 10 tags, and an accuracy test (AT) for 2 tags. The sampling rate was 1 Hz. In SMPT, tags were held for one minute in 12 predefined measuring points, with coordinates known from the barn. In AT, the tags were moved back and forth on eight specific measuring lines (3.45–3.63 m) determined by a wooden plank mounted on the top of the cubicle structures. The coordinates of the measuring lines' endpoints were known. In SMPT every sample points' distance to the true measuring point was calculated using euclidean distance :

$$P_1P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

and in AT distance to the true measuring line:

$$d = \frac{a_{x0} + b_{y0} + c}{\sqrt{a^2 + b^2}}$$

Results and discussion

In SMPT the distance of the sample point and measuring point was 0.17±0.17 m (mean±SD; median 0.16 m; min–max 0.02–2.03 m, see Table 1.). Twenty-seven out of the 465 sample points (6 %) had higher distance than the 30 cm promised accuracy by the manufacturer. In AT the distance to the measuring line was 0.10±0.09 m (mean±SD; median 0.08 m; min–max 0–0.74 m, Table 2.). Sixteen out of 343 sample points (5 %) had higher distance error than that claimed by the manufacturer. Figures 1 and 2 give example of sample points in two measuring lines.

Table 1. Mean, standard deviation, minimum and maximum of the distances (m) between sample points and measuring points

Measuring point	Tag	Mean	SD	Min	Max	The number of sample points
1	051-157	0.19	0.11	0.12	0.47	17
	062-077	0.12	0.01	0.11	0.17	29
2	053-105	0.08	0.01	0.06	0.10	34
	053-119	0.06	0.01	0.04	0.08	24
3	051-157	0.25	0.03	0.17	0.28	35
	062-077	0.26	0.02	0.20	0.29	32
4	053-105	0.16	0.00	0.16	0.17	13
	053-119	0.10	0.01	0.08	0.11	11
5	051-157	0.31	0.39	0.12	1.61	31
	062-077	0.22	0.12	0.12	0.51	9
6	053-105	0.15	0.09	0.07	0.34	9
	053-119	0.20	0.02	0.17	0.25	9
7	053-101	0.10	0.04	0.06	0.14	3
	053-165	0.15	0.23	0.03	0.91	34
8	053-103	0.25	0.01	0.24	0.27	22
	053-159	0.24	0.01	0.22	0.25	15
9	004-190	0.12	0.04	0.07	0.21	18
	053-162	0.19	0.04	0.10	0.25	17
10	004-190	0.51	0.46	0.27	2.04	15
	053-162	0.23	0.03	0.20	0.30	17
11	053-101	0.09	0.02	0.04	0.12	36
12	053-103	0.17	0.01	0.16	0.19	25
	053-159	0.12	0.01	0.11	0.13	10
Total		0.19	0.17	0.03	2.04	465

Table 2. Mean, standard deviation, minimum and maximum of the distances (m) of sample points to the measuring lines

Measuring line	Tag	Mean	SD	Min	Max	The number of sample points
1	017-189	0.07	0.06	0.00	0.16	16
	048-021	0.11	0.09	0.01	0.38	21
2	017-189	0.06	0.04	0.00	0.13	17
	048-021	0.05	0.03	0.01	0.10	16
3	017-189	0.04	0.03	0.00	0.09	22
	048-021	0.02	0.02	0.00	0.07	15
4	017-189	0.07	0.05	0.00	0.20	22
	048-021	0.06	0.05	0.00	0.13	20
5	017-189	0.09	0.04	0.01	0.19	23
	048-021	0.12	0.03	0.07	0.16	23
6	017-189	0.13	0.06	0.07	0.32	26
	048-021	0.09	0.04	0.01	0.19	27
7	017-189	0.09	0.05	0.01	0.16	22
	048-021	0.06	0.04	0.01	0.19	24
8	017-189	0.24	0.18	0.03	0.74	23
	048-021	0.16	0.13	0.01	0.45	26
Total		0.10	0.09	0.00	0.74	343

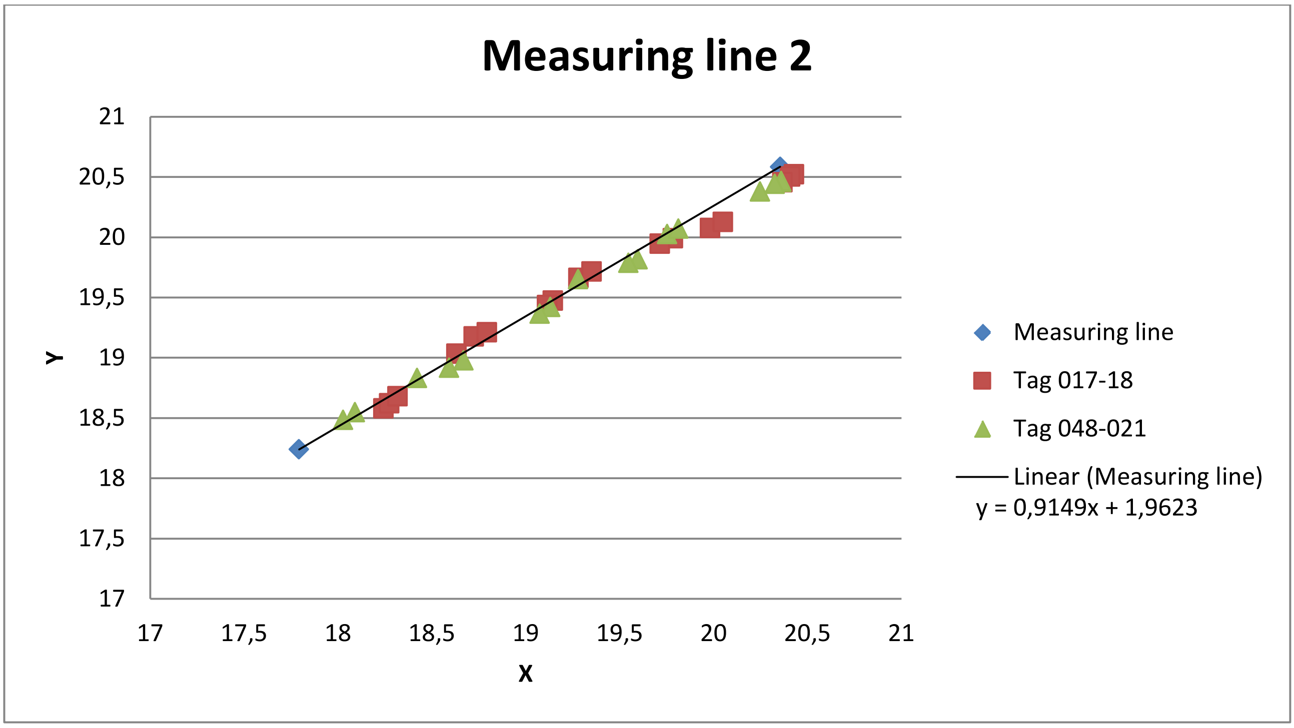


Figure 1. Measured sample points in measuring line 2.

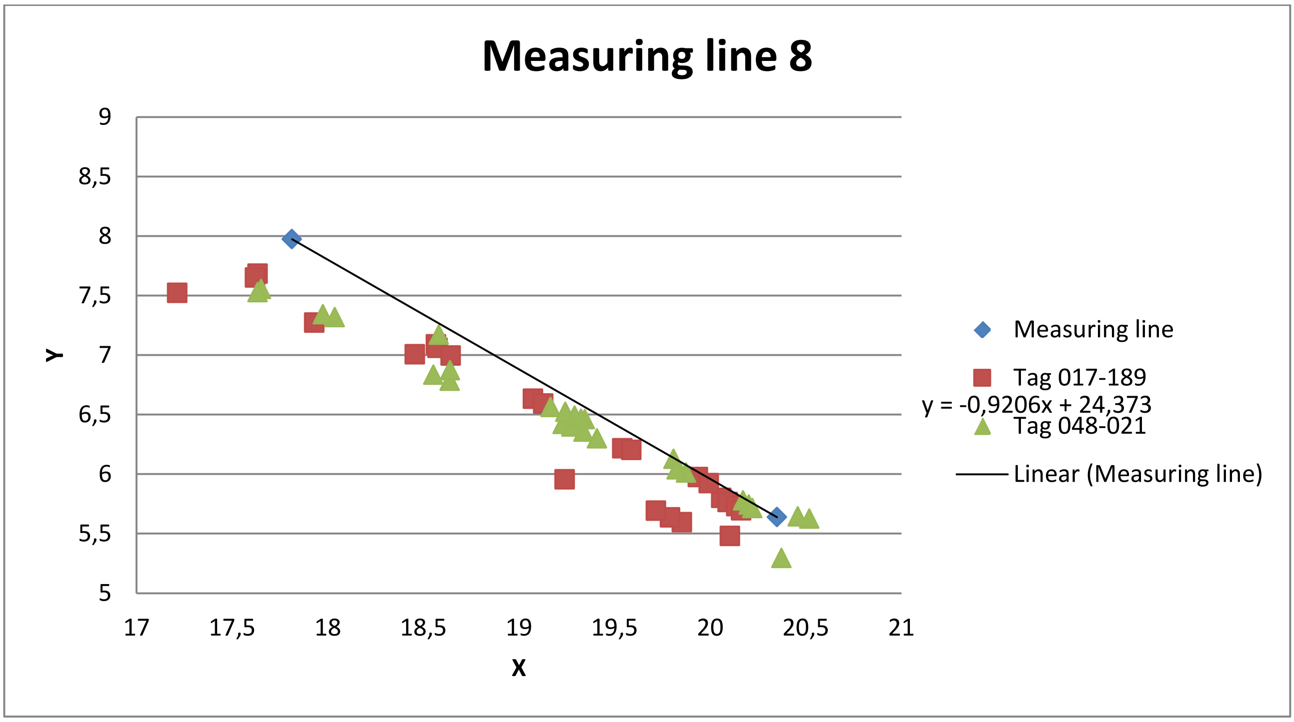


Figure 2. Measured sample points in measuring line 8.

Approximately 95% of sample points were within the promised accuracy. Most of the clearly deviating values seemed to occur when the sensors temporarily lost signal of the tag. At least two out of the six sensors need to get the signal of the tag for a valid positioning. Possible reasons for losing the signal were barn structures. The body of the experimenter could have also blocked the signals. Metal surfaces, e.g. in feeding station, probably reflected the signals and, thus, caused artefacts. In conclusion, TrackLab with Ubisense system is a promising tool for positioning in a cow barn environment. However, more validation tests will be carried out, especially in real life situations with cattle and also different data filtering options will be explored.

References

[1] Homer, E.M., Gao, Y., Meng, X., Dodson, A., Webb, R., Garnsworthy, P.C. (2013). Technical note: A novel approach to the detection of estrus in dairy cows using ultra-wideband technology. *Journal of Dairy Science*, ***96***, 6529-6534.

[2] Gygax, L., Neisen, G., Bollhalder, H. (2007). Accuracy and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. *Computers and Electronics in Agriculture*, ***56***, 23-33

[3] Huhtala, A., Suhonen, K., Mäkelä, P., Hakojärvi, M., Ahokas, J. (2007). Evaluation of instrumentation for cow positioning and tracking indoors. *Biosystems Engineering*, ***96***, 399 – 405.

Modeling Feeding Behavior and Associated Welfare and Productivity of Pigs: an Agent-based Approach

Boumans, I.J.M.M.[1], Hofstede, G.J.[2], De Boer, I.J.M.[1] and Bokkers, E.A.M.[1]

[1] Animal Production Systems group, Wageningen University, Wageningen, the Netherlands
iris.boumans@wur.nl
[2] Information Technology Group, Wageningen University, Wageningen, the Netherlands

Introduction

For animals living in groups such as commercially kept growing pigs, social factors can play an important role in feed intake. Social conflicts around the feeder can change feed intake related behavior of pigs, such as the number of feeder visits, duration of the visits and feeding rate [1]. Conflict situations can increase aggressive behaviors, increase stress, and reduce feed intake. Reduced feed intake in itself can affect growth and health of pigs, increase stress and may result in abnormal behaviors, such as tail biting or oral stereotypies [2]. Altogether, conflict situations and reduced feed intake may impair pig welfare and productivity. The aim of this study was to identify key factors controlling feeding behavior of growing pigs and to examine the effect of social factors on feeding behavior and associated welfare and productivity. Results of this study provide a theoretical framework for a computer simulation (agent-based model), that allows to explore effects of social factors in various situations on feeding behavior of pigs and associated welfare and productivity indicators.

Theoretical framework

A theoretical framework was constructed, based on empirical data and theories in literature, to identify the essential elements affecting feeding behavior of pigs. Behavioral patterns around feeding result from a complex interaction between various mechanisms. To understand how these mechanisms interact, the framework integrates knowledge of ethology, physiology, psychology, and nutrition at animal level and knowledge on the effect of group dynamics and housing conditions on feeding behavior. Many empirical studies show the effect of pig characteristics, physiological factors, nutritional factors, physical factors, and social factors on feeding behavior of pigs [e.g. 1, 3, 4]. Various ethological and psychological theories that explain behavior include psychological factors, such as goals, motivations, emotions and expectations [e.g. 5, 6]. Unfortunately, empirical data on the effect of psychological factors on feeding behavior is limited. To include psychological factors in the theoretical framework, and to explain and understand how internal and external factors affect feeding behavior, theories of motivation are applied in the framework. These theories state that internal factors can affect the response of animals to external factors, and explain that both internal and external factors can motivate an animal to perform a certain behavior.

Agent-based model

Based on the theoretical framework, pig feeding behavior is modeled in an agent-based model. Agent-based modeling (ABM) is a computational modeling method based on heterogeneous agents. Agents are individuals or entities in the simulation with programmed characteristics that can act autonomously and can interact with other agents and the environment. Each agent can adapt and change its behavior during the simulation, allowing for unpredicted and complex behavioral patterns to emerge at a higher level. ABM is a suitable approach for addressing complex and multilevel systems and gaining insight in the interaction of the individual components and the effect on system behavior [7].

In this study, a two dimensional and spatially explicit agent-based model was constructed and implemented in Netlogo 5.0.3 [8]. The environment in the model represents a conventional barren housing pen for group housed pigs with one feeder and ad libitum access to feed. The agents represent growing pigs from the age of 10 till 25 weeks. Variation in pig characteristics can be modeled, for example, by assigning pigs different growth capacities (associated with genotypes) or different coping styles (associated with active or passive stress responses). Dynamics of feeding behavior are modeled by combining a nutritional and physiological growth model [9, 10] with a mechanistic control of feed intake and physiological feedback mechanisms. The control of feed intake is affected by five factors: pig characteristics, feed characteristics, physiological factors, psychological factors, and social factors (Figure 1). The theoretical framework is implemented in the model as two processes within each agent. The first process concerns the formation of feeding motivation and other motivational states in the pigs, the second process is directed at decision-making, i.e. how pigs act behaviorally on the feeding motivation. The processes in the simulation are updated every time step (representing one minute).

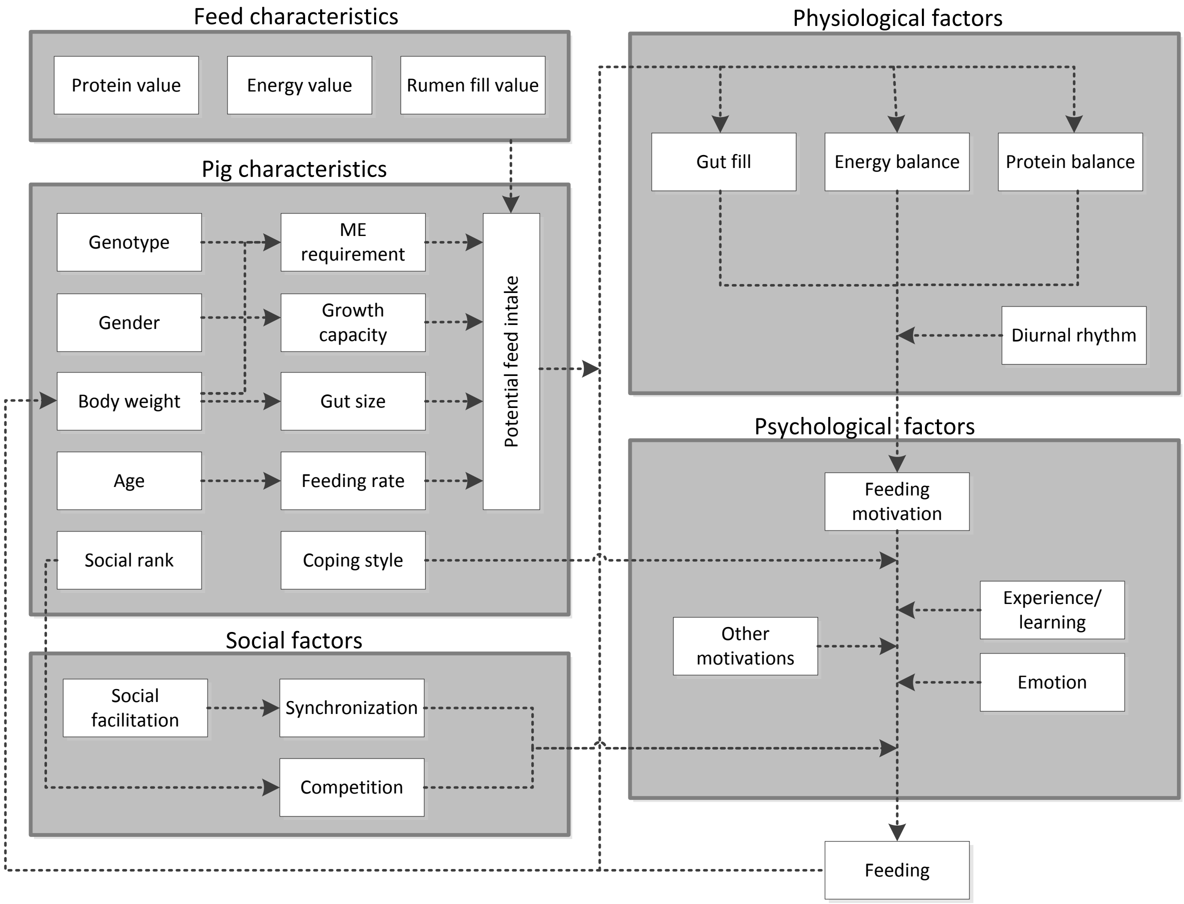


Figure 1. Schematic overview of feeding behavior in the model: the formation of feeding motivation and factors affecting decision-making of pigs to act on feeding motivation.

Model output includes different behavioral components of feeding behavior, such as amount of feed intake, number of visits, meal duration and intervals between meals. Furthermore, output also includes the sequence of behaviors and time budget of pigs (for performing certain behaviors). Behaviors that can precede or follow feeding behavior (e.g. waiting, drinking, or approach) are included in the model to allow pigs to adapt to conflict situations. Adaptation of pigs to conflict situations is affected by coping style, competition (influenced by social rank), and psychological factors, such as experience or motivations. The model allows to monitor the development of pigs and possible changes in their behavior in time. Furthermore, several welfare and productivity indicators can be assessed, such as number of conflicts, aggressive encounters, stress levels and feed efficiency of pigs (based on feed intake, energy use, and growth).

Conclusion and future work

The interdisciplinary theoretical framework shows how individual variation among pigs and social influences interact multidimensionally and can affect feeding behavior. ABM allows to test theoretical models and to gain insight in causation and consequences of behavior. The agent-based model is developed in two steps to reduce complexity of the model and to understand the effect of internal and external factors separately. The first step is the development of a model that reflects the dynamics of feeding behavior affected by internal factors, such as physiological factors, psychological factors and pig characteristics (by modeling an individual housed pig). The second step concerns the inclusion of social factors and the effect of interaction among pigs (and modeling group housed pigs). Examples of questions that can be studied with this model are: (1) Which internal factors can explain behavioral feeding patterns of a pig?; (2) How can social factors affect feeding behavior and welfare?; (3) How can pig characteristics and group composition affect feeding behavior and welfare? The model in this study is still preliminary and needs to be further refined and evaluated before final conclusions can be drawn. First results show that diurnal behavioral patterns can be modeled in which pigs are more active and perform more feeding behavior during daytime than during the night. In the future, the model will be validated by aligning the model to experimental settings in available experimental data and studies described in the literature, and by comparing results of these studies with the output of the model.

References

[1] Bornett, H.L.I., Morgan, C.A., Lawrence, A.B., Mann, J. (2000). The effect of group housing on feeding patterns and social behaviour of previously individually housed growing pigs. *Applied Animal Behaviour Science, **70**(2)*, 127-141.

[2] Meunier-Salaün, M.C., Edwards, S.A., Robert, S. (2001). Effect of dietary fibre on the behaviour and health of the restricted fed sow. *Animal Feed Science and Technology, **90**(1–2)*, 53-69.

[3] Quiniou, N., Dubois, S., Noblet, J. (2000). Voluntary feed intake and feeding behaviour of group-housed growing pigs are affected by ambient temperature and body weight. *Livestock Production Science, **63**(3)*, 245-253.

[4] Nielsen, B.L., Lawrence, A.B., Whittemore, C.T. (1995). Effect of group size on feeding behaviour, social behaviour, and performance of growing pigs using single-space feeders. *Livestock Production Science, **44**(1)*, 73-85.

[5] Jensen, P., Toates, F.M. (1997). Stress as a state of motivational systems. *Applied Animal Behaviour Science, **53**(1–2)*, 145-156.

[6] Toates, F. (2004). Cognition, motivation, emotion and action: a dynamic and vulnerable interdependence. *Applied Animal Behaviour Science, **86**(3–4)*, 173-204.

[7] Railsback, S.F., Grimm, V. (2012). *Agent-based and individual-based modeling: a practical introduction*. Vol. Princeton, New Jersey. Princeton University Press.

[8] Wilensky, U. (1999). Netlogo. <http://ccl.northwestern.edu/netlogo/>. Accessed 18 November 2013.

[9] de Lange, C.F.M. (1995) *Framework for a simplified model to demonstrate principles of nutrient partitioning for growth in the pig*, in *Modelling growth in the pig.*, Moughan, P.J., Verstegen, M.W.A., and Visser-Reyneveld, M.I., Editors. Wageningen Press: Wageningen, Netherlands. 71-85.

[10] National Research Council (2012). *Nutrient requirements of swine*. Animal nutrition series. Washington, D.C: National Academies Press.

Tracking individual laying hens in group housing

T. B. Rodenburg, M. Naguib

[1]Behavioural Ecology Group, Wageningen University, Wageningen, The Netherlands
bas.rodenburg@wur.nl

Introduction

Many animal species live in social groups and show complex social behaviour. Social behaviour has been well studied on various levels, ranging from classical ethological approaches, i.e. behavioural observations, to experimental and evolutionary approaches focussing on the fitness value of group living. Yet while research on smaller groups has provided important insights into behavioural mechanisms, such research on larger groups has been hampered by the inability to keep track of the movement and behaviour of individuals. For large populations, predominantly approaches have been taken that describe group or flock behaviour, for instance swarming behaviour in birds [1], schooling behaviour in fish [2] and social dynamics in humans [3], but those approaches usually do not allow linking individual characteristics to group behaviour. In smaller groups, an individual-based approach allows us to link the characteristics of a given individual to the behaviour of the group. However, challenges arise when these methods are applied to large groups of animals due to challenges associated with a larger number of animals interacting with each other and very different social dynamics.

Tracking individuals in groups

To be able to study the effects of individual variation on behaviour of large groups, it is essential to bridge the gap between individual behaviour and group behaviour. This is specifically important in husbandry systems, where animals are kept in large groups and where "unwanted" behavioural traits, like damaging behaviour, can spread like an epidemic throughout the group. The laying hen provides an excellent model to study this: in commercial practice, laying hens are kept in very large groups where almost nothing is known on individual behaviours. To date, various approaches have been taken to study the behaviour of individual laying hens in group housing. Collins [4] used video tracking to track individual broilers. One disadvantage of video tracking, however, is that it is very difficult to identify and keep track of your focal animal, especially in large groups. A second possibility is to use RFID technology, with PIT tags on the legs of the animals that are read when birds pass an antenna. The success of this approach is mainly linked to the correct placement of the antennas. Kjaer [5] formed a grid of antennas and used this set-up to show differences in general activity levels between birds from high and low feather pecking lines, linking feather pecking to hyperactivity. RFID technology can also be very useful to study use of an outdoor run, where birds have to pass a narrow pop-hole in order to get access to that area. By placing antennas in the pop-holes, you can get a good idea of individual use of the outdoor area [6]. Using an approach like this, we are currently investigating how outdoor usage is influenced by initial location in the poultry house (near or far from the pop-holes) and by other behavioural characteristics of the individual bird (for instance personality traits).

Sensor technology

Probably the best opportunities for tracking individual hens in groups are offered by sensor technology, using active sensors. Using this method, location of individuals can be monitored continuously and also be combined with accelerometer and proximity data. An important aspect of this type of tracking is that the tags should not change the behaviour of the focal animal or make the focal animal more or less attractive for social interactions. Daigle et al. [7] showed that no effects on behaviour were found when using lightweight (10 g) sensors. Using the same system, Quwaider et al. [8] showed that location tracking using the sensors was for 84% similar to the results yielded by video tracking the same birds. Furthermore, they showed that the accelerometer data could be used to automatically quantify specific behaviours, by linking accelerometer data to video data. Another promising tracking method is proximity logging: when you want to understand social behaviour of a group of animals, it is a very powerful tool to be able to record social interactions between individuals in that group. Further, if you think of damaging behavior, such as feather pecking, proximity data could be very useful to identify the culprits: i.e. the bird that was most frequently near the birds with severe feather damage. To be able to collect this kind of data, we will use the Encounternet® system, a system our group also uses to track wild song-birds, in group-housed laying hens. This system combines the ability to detect bird location with accelerometer and proximity data.

Conclusion

Sensor technology provides powerful tools to investigate behaviour of individual laying hens in group housing systems. These approaches will allow us to better understand social behaviour of laying and to investigate the effects of genetic and early-life manipulations on social behaviour. Further, more robust methodology, such as RFID technology, could also be used to investigate social behaviour and use of facilities on commercial farms. Together, this type of methodology should help us to provide a better understanding of the role of individual animals in large groups and may also offer useful tools to identify individuals that negatively affect group performance. Such tools could for instance be used by breeding companies, when selecting the birds for the next generation.

References

[1] Hildenbrandt, H., Carere, C., Hemelrijk, C.K. (2010). Self-organized aerial displays of thousands of starlings: a model. *Behavioral Ecology,* ***21****(6)*, 1349-1359.

[2] Hemelrijk, C.K. et al. (2010). Emergence of oblong school shape: Models and empirical data of fish. *Ethology,* ***116****(11)*, 1099-1112.

[3] Krause, J., Ruxton, G.D., Krause, S. (2010). Swarm intelligence in animals and humans. *Trends in Ecology & Evolution,* ***25****(1),* 28-34.

[4] Collins, L.M. (2008). Non-intrusive tracking of commercial broiler chickens in situ at different stocking densities. *Applied Animal Behaviour Science,* ***112****(1-2)*, 94-105.

[5] Kjaer, J.B. (2009). Feather pecking in domestic fowl is genetically related to locomotor activity levels: implications for a hyperactivity model of feather pecking. *Behavior Genetics,* ***39***, 564-570.

[6] Richards, G.J. et al., (2011).Continuous monitoring of pop hole usage by commercially housed free-range hens throughout the production cycle. *Veterinary Record,* ***169****(13)*, 338-358.

[7] Daigle, C.L. et al. (2012). Noncaged laying hens remain unflappable while wearing body-mounted sensors: Levels of agonistic behaviors remain unchanged and resource use is not reduced after habituation. *Poultry Science,* ***91****(10)*, 2415-2423.

[8] Quwaider, M.Q. et al. (2010). Development of a wireless body-mounted sensor to monitor location and activity of laying hens in a non-cage housing system. *Transactions of the ASABE,* ***53****(5)*, 1705-1713.

Introduction to the symposium on Ambulatory Skin Conductance

Martin Ouwerkerk and Joyce Westerink

Philips Group Innovation – Research

Ambulatory skin conductance as we see it now emerging on the market with products like the Affectiva Q-sensor and the Basis B1, was promoted already in 1997 by Picard and Healey [1]. Also prototype devices such as the Empatica E3, the Smartband, and the Philips DTI-2 ([3] and figure 1) typically measure skin conductance at the base of the wrist, since this is a convenient measurement position. However, traditionally skin conductance (electrodermal activity, or galvanic skin response) is measured on the fingers, the palmar region of the palms, or on the soles of the feet. A traditional skin conductance measurement usually is also corrected with a baseline measurement, while for ambulatory measurements this needs to be self-calibrating. These differences in measurement position and procedure are among the factors that cause differences in measurement results.

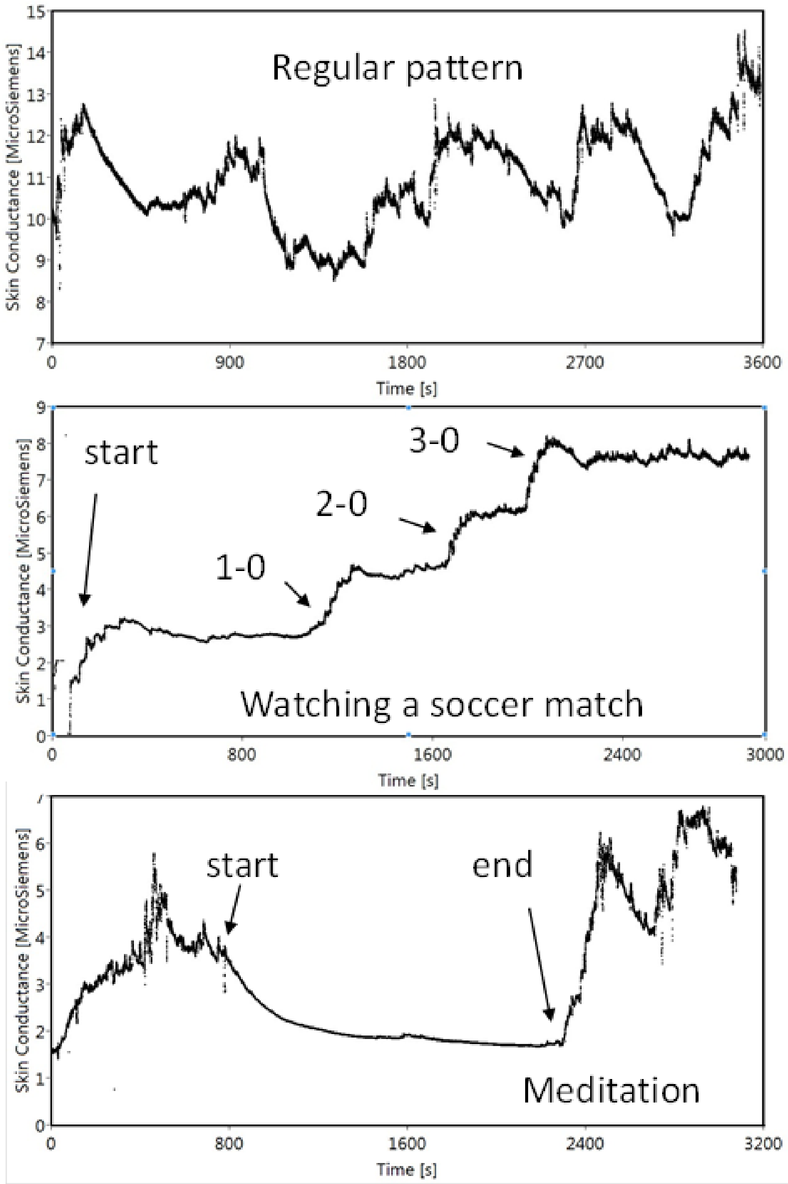


Figure 1. Skin conductance patterns obtained with the Philips DTI-2 prototype sensor wristband.

A comparison of skin conductance at the base of the wrist and on the fingers was published by Poh et al . (2010) [2]. A high accuracy and a strong correlation is claimed. In a study by van Dooren et al. (2012) [4] no less than 16 positions were compared. Fingers, foot, and forehead clearly offer the highest skin conductance with strongest skin conductance responses. At the wrist the responses are weaker but still present. On other body locations almost all skin conductance responses were absent. Apparently differences between measurement locations exist (see figure 2), although it is not clear whether these differences preclude the user of trustworthy ambulatory (wrist) measurement positions.

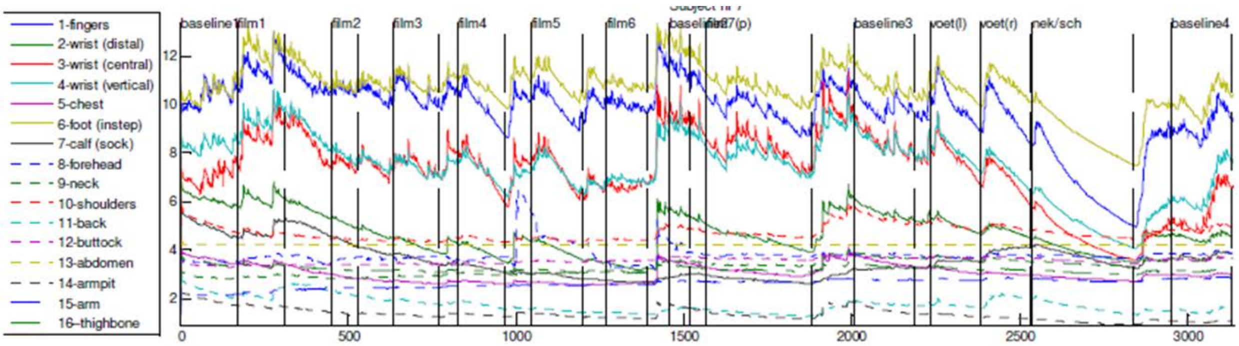


Figure 2. Skin conductance measured at 16 body positions simultaneously while subjecting the participant to a visual stimulus.

Another peculiar skin conductance phenomenon is the occurrence of non-responders and hyperresponders (see figure 3). Boucsein addresses it in his book (2012) [5]. Especially with psychopaths and schizophrenics non-typical electrodermal responses have been reported. Both non-responders and hyperresponders are sometimes removed from laboratory measurements, but have to be dealt with in practical applications.

This symposium addresses the applicability of ambulatory skin conductance measurement for practical applications. Are the measurements trustworthy enough to warrant useful applications, even when not measured at the traditional measurement positions? Is it possible to do formal baseline measurements in practical applications, and if not, is that a problem? How often do we find non-responders or hyperresponders in the populations of our interest?

The 5 contributors to the Ambulatory symposium each address these questions from a different angle. On the one hand, Eco de Geus presents data that shows that even properly baselined ambulatory skin conductance measurements should be supported by other measurements to give a true impression of the activity of our sympathetic nervous system. Nevertheless, the other symposium contributors explore the possibility of practical applications. Rafal Kocielnik describes a system that helps employees to get a better impression of their daily stress patterns, whereas Matthijs Noordzij explores the use of ambulatory skin conductance in the care for severely mentally disabled patients. Ewout Meijer discusses new developments in the old science of detection of deception, for instance expanding the domain to group measurements, which requires new measurement methodologies. Finally, Henk Nijman focuses on extracting useful information from skin conductance measured on patients in psychiatric wards, like indicators of upcoming aggression.

Overall we hope that this symposium gives a balanced overview of trustworthiness of ambulatory skin conductance measurement in practical applications.

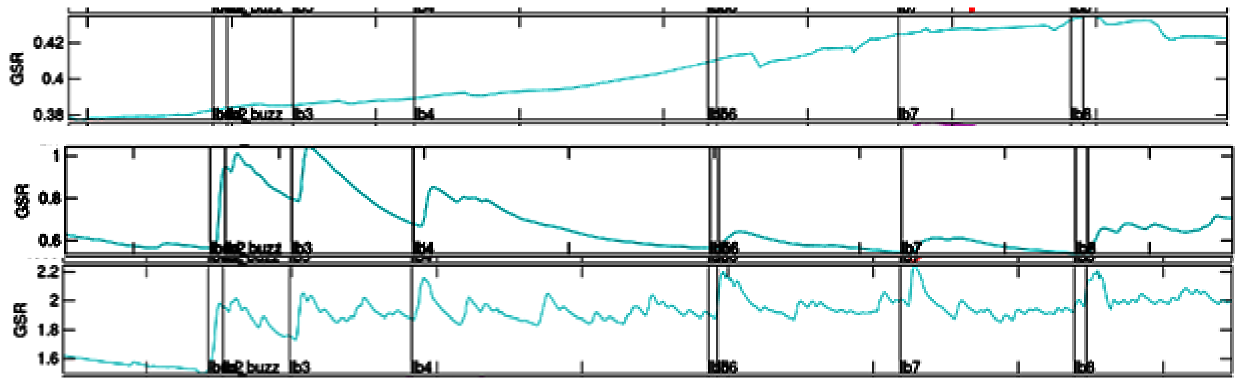


Figure 3. Skin conductance trace measured at the fingers during a number of loud buzzes (lb1 until lb8) : a) non-responder, b) normal response dying out, and c) hyperresponder.

References

[1] Picard, R.W., Healey, J. (1997). Affective wearables. *Personal Technologies, **1**(4)*, 231-240.

[2] Poh, M.-Z., Swenson, N.C., Picard, R.C. (2010). Wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, ***57**(5)*, 1243 – 1252.

[3] Ouwerkerk, M., Dandine, P., Bolio, D., Kocielnik, R., Mercurio, J., Huijgen, H., Westerink, J. (2013). Wireless multi sensor bracelet with discreet feedback. In *Proceedings of the 4th Conference on Wireless Health (WH'13)*, 1–3 November, Baltimore, MD, USA. Article No. A6. New York, NY, USA: ACM.

[4] van Dooren, M., de Vries, J.J.G., Janssen, J.H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, ***106**(2)*, 298–304.

[5] Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). New York, NY, USA: Springer Science+Business Media, LLC.

Comparing cardiac and skin sympathetic nervous system activity

Eco J.C. de Geus, Annebet Gerssen-Goedhart, Gonneke Willemsen

Department of Biological Psychology, VU University, Amsterdam

Abstract

The Vrije Universiteit Ambulatory monitoring system can in principle measure the pre-ejection period (PEP) from the impedance cardiogram and skin conductance level (SCL) form two palmar electrodes in ambulatory real life settings. To test whether PEP and SCL quantify sympathetic nervous system (SNS) activity in a comparable way data were first obtained under standardized lab conditions from 39 human subjects during exposure to 13 different experimental conditions. Compared to pre-test resting baseline recordings significant decreases in PEP and parallel increases in the SCL were found for various stressors known to increase SNS activity. The between and within subjects correlations between PEP and SCL were not significant. This suggests that SNS activity is reflected differently by the heart and the skin. We conclude that SNS activity studies, when possible, should include both PEP and SCL measurement.

Introduction

The Vrije Universiteit Ambulatory Monitoring System (VU-AMS) can measure the pre-ejection period (PEP) from the impedance cardiogram and skin conductance level (SCL) from two palmar electrodes in ambulatory real life settings. In principle this would allow us to measure sympathetic nervous system (SNS) activity from two entirely different organ systems. Under the assumption that SNS activity to most mental and physical stressors presents a unitary response, this dual read-out of PEP and SCL would provide us with a convergent and robust estimation of SNS activity. However it has been questioned whether SNS activity indeed presents itself as a unitary response across multiple organ systems. In this study we investigated the extent to which Pre-Ejection Period (PEP) and Skin Conductance Level (SCL) responses are correlated, and whether they can be considered to measure SNS activity in a comparable way.

Methods

Physiological data were obtained from 38 human subjects (23 males) during 8 different mental or physical stressors and 5 resting/recovery periods. The ECG and ICG were recorded using seven pregelled Ag/AgCl spot electrodes (UltraTrace, ConMed, USA) in the typical VU-AMS configuration shown in Figure 1. The electrodes were connected to the ECG100C and NICO100C BioPac modules using extension leads.

Skin Conductance information was collected using a pair of Ag/AgCl (unpolarizable) electrodes (Ø = 6 mm) attached with a Velcro strap to the distal phalanx of the index and middle finger of the non-dominant hand. To ensure sufficient electrode–skin contact, isotonic electrode paste was used (0.5% saline in a neutral base). The electrodes were connected to the GSR100C BioPac module.

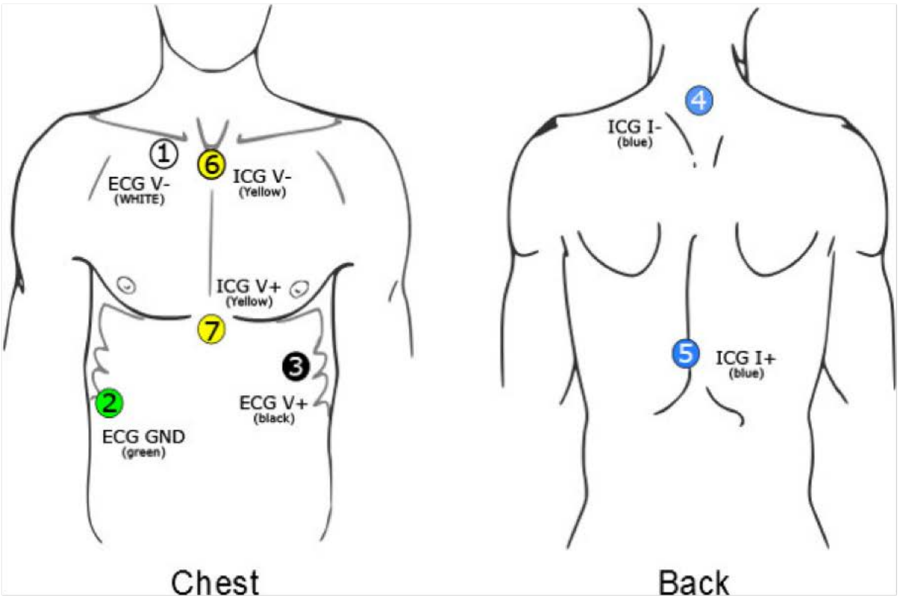


Figure 1. The seven electrodes should be placed on the participant's chest and back. The first ECG electrode (V-) is placed slightly below the right collar bone 4 cm to the right of the sternum. The second ECG electrode (V+) is placed at the apex of the heart over the ninth rib on the left lateral margin of the chest approximately at the level of the processus xiphodius. The third ECG electrode (GND) is a ground electrode and is placed on the right side, between the lower two ribs at the right abdomen. The first ICG measuring electrode (V_1) is placed at the top end of the sternum, between the tips of the collar bones. The second ICG measuring electrode is placed at the xiphoid complex of the sternum, where the ribs meet. The two current electrodes are placed on the back: I- on the spine over the cervical vertebra C4, at least 3 cm (1 in) above the ICG measuring elec-trode V-, and I+ between thoracic vertebrae T8 and T9 on the spine, at least 3 cm (1") below the ICG measuring elec-trode V_2. The ICG electrode placement takes into account that the largest part of the left ventricle driven change in thorax impedance is captured by the column between the suprasternal notch and the processus xiphoideus.

The various experimental conditions were explained to the subject and the mental and physical stress tasks were briefly practiced. The actual experiment started by asking the subjects to sit quietly and relax for a pre-test 10 minutes resting baseline. Next, the following conditions were presented in a fixed order: Stroop colour word task (4 min), recovery1 (3 min), tone avoidance task (4 min), recovery2 (3 min), lying (2 min), standing (2 min), recovery3 (2 min), hand grip test (2 min), cold pressure test (1 min), recovery4 (3 min), and the step test (2 min). After the step test a final post-stress resting condition of 13 minutes concluded the physiological recordings.

Pearson correlations were computed between PEP and SCL separately for each of the conditions (between-subject correlations) and separately for each of the subjects across all conditions (within-subject correlations).

Results

Between-subject correlations show PEP to be largely uncorrelated with SCL. Within-subject correlations across the 13 experimental conditions are shown in Table 3 (modified from Goedhart AD, Willemsen G, De Geus EJC (2008). *Sympathetic nervous system activity in the heart and the skin: are they comparable?* In M. Kaneko (Ed.), Sympathetic Nervous System Research Developments. New York: Nova Science Publishers. pp93-114). The mean within-subject correlation between PEP and SCL was -.43, but large individual differences were found. Furthermore, inspection of the scatter plots strongly suggested that the relation between PEP and SCL entirely depended on the exercise condition. When we recomputed the within-subject correlations after exclusion of the step test data, the correlation between PEP and SCL remained significant in one subject only.

Conclusion

We show that PEP and skin conductance respond to stress tasks in a manner compatible with increased SNS activity but that within the same subject their response is largely uncorrelated. The heart and the skin, therefore, seem to reflect different aspects of SNS activity. Since both measures have already shown their usefulness in biomedical and psychophysiological research, PEP and SCL should ideally be measured in parallel.

Table 1. Within-subject correlations between PEP and SCL for all conditions, and for all conditions except the step test.

	All conditions		*All conditions except the step test condition*	
Subject	$r_{\text{PEP-SCL}}$	p	$r_{\text{PEP-SCL}}$	p
1	-.11	.72	.46	.14
3	.51	.09	.51	.09
5	**-.64**	.02	-.03	.94
6	-.54	.06	.45	.14
7	-.47	.11	.32	.31
8	**-.62**	.02	-.27	.39
9	-.29	.35	-.29	.35
10	**-.73**	.00	-.16	.72
11	**-.70**	.01	-.48	.11
12	-.44	.13	-.22	.49
13	**-.56**	.05	-.10	.76
14	**-.71**	.01	**-.73**	.01
15	-.46	.11	.25	.44
16	-.35	.24	.00	1.00
17	**-.65**	.02	-.09	.79
18	-.38	.20	.12	.72
19	-.36	.23	-.01	.98
20	**-.72**	.01	-.16	.62
21	-.23	.46	.25	.44
22	-.43	.15	.14	.66
23	**-.66**	.01	-.18	.58
24	**-.81**	.00	-.21	.52
25	**-.73**	.00	-.19	.55
26	**-.66**	.01	-.05	.88
28	.41	.18	.41	.18
29	-.46	.11	-.42	.18
30	-.15	.63	.26	.42
31	-.29	.34	.01	.98
32	.49	.09	-.22	.50
33	-.08	.80	.44	.16
34	**-.79**	.00	.01	.98
35	**-.77**	.00	.19	.56
36	**-.59**	.03	-.05	.88
41	-.41	.19	-.41	.19
42	-.55	.05	.35	.27
43	-.02	.96	.25	.44
45	**-.61**	.03	-.01	.99
47	**-.84**	.00	-.37	.24
48	-.49	.09	-.04	.90

Bold: significant at $p < .05$ level.

The Use of Skin Conductance in the Detection of Deception

E.H. Meijer

Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands
Eh.meijer@maastrichtuniversity.nl

Historically, measures of the Autonomic Nervous System (ANS measures) have played a prominent role in both research on, and application of deception detection procedures. Most diagnostic of these measures is the Skin Conductance Response [1], which was first introduced to the field of deception detection in 1939 [2].

But despite their longstanding application, the use of skin conductance in deception detection is highly controversial [3]. Much of this criticism, however, is not so much geared towards the shortcomings of measures such as skin conductance, but rather to the Control Questioning Test (CQT) format it is used in conjunction with. In the CQT test format, ANS responses to relevant questions (e.g., Did you kill X) are compared to responses to control questions (e.g., Did you ever do anything illegal). Stronger responses to the relevant than to the control questions leads to a 'deception indicated' outcome, while stronger responses to the control questions lead to a 'no deception indicated' outcome. Relevant and control questions, are however, confounded with a variety of psychological processes, including emotional valence, making any conclusion based on the comparison cumbersome.

Because of the shortcomings of the CQT, several other questioning formats were developed. Among these is the Concealed Information Test (CIT; [4]), which has most extensively been used with skin conductance response as the dependent measure. In the CIT, test questions address crime related details that are known only to the perpetrator and the investigative authorities. For example, a test question may address the location where a body was found, and this question is presented with several alternatives (e.g., a. in a bedroom; b. in a pond; c. in a driveway; d. in a garage; e. in an ally). To an innocent suspect these alternatives are equally plausible, and will therefore evoke similar skin conductance responses. For a guilty suspect, the correct alternative stands out, and will therefore evoke a larger response. Thus, guilt is inferred if a suspect systematically shows an enhanced response to the correct alternative.

Whereas research on the (the development) of the CQT has come to a virtual standstill, research on the CIT is still actively pursued by researchers worldwide. This includes several new challenges including: 1) the extent to which the CIT can be used when the correct alternative is unknown – and is the topic of investigation (e.g., [5]); and 2) the use of the CIT in groups (e.g., groups of terrorism suspects) rather than individuals; (e.g., [5,6]. These new applications require differ methodologies than typically used the field of deception detection. This includes differences in the way skin conductance data is collected, as well as differences in the algorithms used for the data analysis. In this presentation, I will highlight two recent studies in which participants were exposed to a mock terrorism scenario (both studies were approved by the standing ethical committee), and the CIT was used to extract information about this mock attack from these participants.

In the first study [6], participants were invited to the lab in groups of 5. Each group was asked to plan a mock terrorist attack based on a list of potential countries, cities, and streets. Next, three questions referring to the country, city, and street were presented, each with five options. Skin conductance in all five members of the group was measured simultaneously during this presentation. A dynamic questioning approach entailed direct analysis of the data, and if the average skin conductance response of the group to a certain option exceeded a threshold, this option was followed up, e.g., if the reaction to the option "Italy" exceeded the threshold, this was followed up by presenting five cities in Italy. Even though effect sizes were only moderate, these results indicate that our dynamic questioning approach can help to unveil plans about a mock terrorist attack.

The second study addressed the research question whether the CIT rationale can also be applied using a continuous recording of skin conductance, rather than skin conductance responses. Participants received information about the location of an upcoming ambush on a money car on route from Jerusalem to Tel Aviv.

Next, the participant was shown an animation of this route on a map, which was repeated four times. During this animation, skin conductance was continuously recorded, and later offline analysed in an attempt to pinpoint the location of the ambush. Results showed some modest support for being able to pinpoint the location, but only using the first repetition of the animation.

Besides the two experiments, I will also address the problems and opportunities associated with recordings of skin conductance in the detection of deception, including which problems can, and which problems cannot be solved by technical developments.

References

[1] Kircher, J.C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, **73***, 291-302.

[2] Grubin, D., & Madsen, L. (2005) Lie detection and the polygraph: A historical review. *Journal of Forensic Psychiatry & Psychology, **16***, 357-369.

[3] National Research Council (2003). *The Polygraph and Lie Detection*. Committee to review the scientific evidence on the Polygraph. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[4] Verschuere, B., Ben-Shakhar, G. & Meijer, E (Eds.). (2011). *Memory detection: theory and application of the Concealed Information Test*. Cambridge University Press.

[5] Meijer, E.H., Smulders, F. & Merckelbach, H. (2010). Extracting concealed information from groups. *Journal of Forensic Sciences, **55***, 1607-1609.

[6] Meijer, E.H., Bente, G., Ben-Shakhar, G., & Schumacher, A. (2013) Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback. *Frontiers in Psychology, **4***, 68.

LifelogExplorer: A Tool for Visual Exploration of Ambulatory Skin Conductance Measurements in Context

R.D. Kocielnik

Department of Mathematics & Computer Science, Eindhoven University of Technology, The Netherlands. r.d.kocielnik@tue.edu.nl

Abstract

We present a user-friendly tool for automated analysis, interpretation and visualization of long-term skin conductance measurements in relation to real life context - LifelogExplorer. The automated processing of raw signal defines personalized stress levels that coupled with powerful interactive visualization offers users a friendly and intuitive way of understanding their physiological reactions in various contexts, such as meetings, hours of the day, etc. Our tool cooperates with the Philips Discrete Tension Indications (DTI-2) measurement wristband, and can easily be extended to use also other measurement equipment. We used the LifelogExplorer in a study with 21 vocational school teachers confirming its usefulness for non-experts.

Introduction

Measuring physiological reactions of the body in real life settings is a big challenge; aside from technical constraints, the challenge lies in discovering and communicating long term physiological patterns to a non-expert user. In this work we focus specifically on measurements related to emotional stress. Stress can lead to diseases such as depression and burnout [1] and indirectly influence development of many other health problems [2]. Use of skin conductance (SC) measurements, as indicator of stress, has been well established in lab settings, where everything is closely monitored and controlled by a researcher. However, the real value of such monitoring lies in performing measurements of stress from regular people and in real-life context when most of the stress is really experienced. Modern wearable sensors begin to offer easy monitoring of various aspects of everyday life continuously and unobtrusively. Usually, however, they don't come with any easy to use software that would offer powerful and understandable interpretations of measurement data for regular users, especially in terms of learning long-term behavioral patterns. External analysis tools such as Matlab and LabView suffer from the same limitation- they are meant for experts. Consequently these devices and the measurements they offer are still mainly used by researchers and "hardcore" enthusiasts from Quantified Self communities.

We offer a tool – LifelogExplorer - that brings the power of real-life SC measurements for estimating the level of stress to the regular user. Our tool automatically processes ambulatory measurements of SC, offers interpretation in form of personalized stress levels and a visual exploration revealing stress patterns in real-life context, all in a user friendly, accessible and simple form. To make the user experience complete we offer easy coupling with an unobtrusive wearable wristband – Discrete Tension Indicator (DTI-2) developed by Philips Research. Our main purpose is to let non-experts understand their stress and reflect on it. Practical usefulness of our tool has been demonstrated in a number of studies in different settings. In this abstract we refer only to the most recent one with 21 vocational school teachers.

The rest of this abstract is structured as follows: (1) we describe our tool – LifelogExplorer - and its main functionalities; (2) we summarize results from our user study, and finally (3) we close with conclusions and future work.

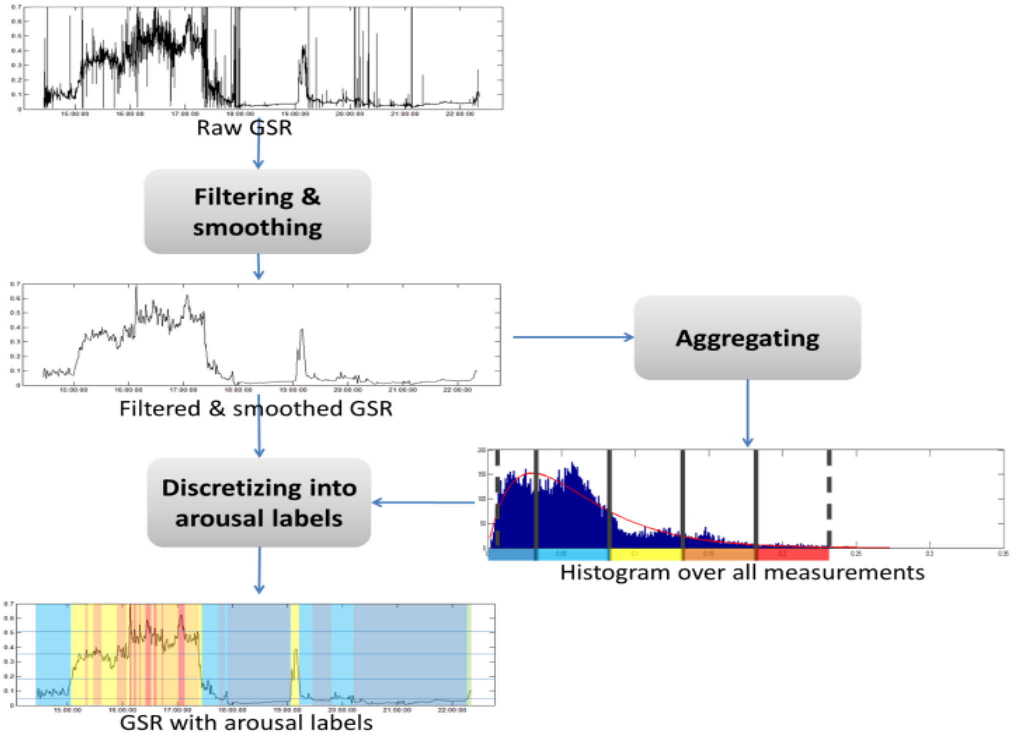


Figure 1. Steps in processing the raw SC measurements obtained from the DTI-2 measurement wristband.

LifelogExplorer

Our goal is to offer integrated systems that automatically obtains physiological measurements from a wearable sensors device, analyzes and presents them to a non-expert using simple and informative interactive visualization. To support the user in learning long-term patterns, our visualization summarizes the measurement data for various practical contexts, such as meeting subjects, participants, locations and such. To meet this goal, the LifelogExplorer tool extends the work done in [3] .We describe selected functionalities of the system below.

Data analysis

The tool cooperates with the Discrete Tension Indicator (DTI-2) measurement wristband created by Philips Research [4] for continuous unobtrusive monitoring of physiological signals and environmental conditions (skin conductance, skin temperature, ambient temperature and lightings, 3D acceleration), allowing to estimate experienced stress level. SC measurements obtained from the DTI-2 are automatically filtered and processed as depicted in Figure 1. A collection of measurements over long term builds up a personalized histogram of physiological responses. This histogram is further used for defining the personalized stress levels represented using a color scale – from blue meaning “calm” to “red” meaning stressed. More precise description of the procedure used can be found in [5].

Calendar views

We use the context of calendar, which most people often work with, as a basis for visualizing short term stress. We offer standard options for defining different time periods being shown: a day, work week, full week. We add a possibility of selecting a set of custom days. This is meant to allow people explore their short term stress in a familiar way. In order to allow people to spot acute stress responses we provide a shape based visualization as shown in Figure 2. The shape visualizes the processed SC measurements in the context of a day. In order to support people in making quick judgments about their day or week we provide a color based visualization as shown in Figure 3. This visualization uses the personalized stress levels for coloring different parts of the day. Through the calendar view we allow the users to explore their short-term stress responses, long-term build-up of stress during the day, and also explore these reactions in different contexts, such as getting stressed during a long, unproductive meeting. We add real-life context to the measurements by importing calendar entries from Outlook, Excel or other source. Aside from these automatically obtained entries we allow users to freely make their own entries if they are interested in tacking or comparing them to the other contexts. Each calendar entry also allows users to indicate their subjective feedback related to the experienced emotions using a questionnaire.

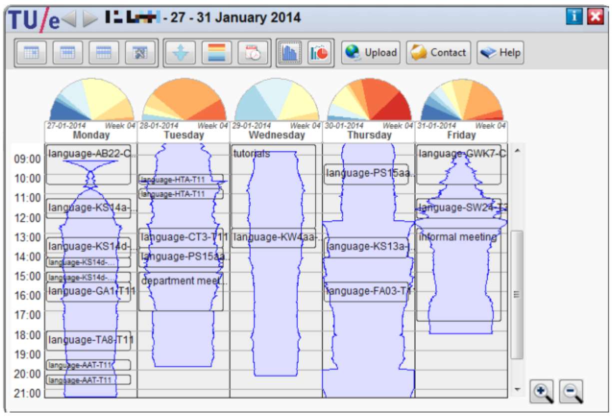


Figure 2. Calendar view with SC measurements represented as shapes in context.

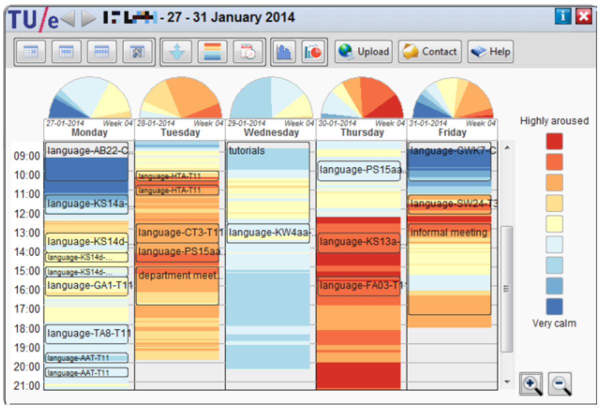


Figure 3. Calendar view with personalized stress levels represented using a color scale from blue - calm to red - stressed.

Aggregated views

To support users in gaining insight into the contexts in which they experience stress repeatedly and consistently, and therefore spot possible stress patterns, we provide a set of aggregated views on the data. Aggregated views summarize the long-term physiological responses in different contexts, such as meeting subject, location, attendees, as well as, days of week, weeks, months and such. Each view is composed of a number of aggregations of stress, each representing different value for a particular dimension.

Figure 4. Aggregated view on "hours of the day", each individual half-pie chart represents a summary of a particular hour across all the measurement days

Figure 5. Aggregated view on "subject of meetings", each individual half-pie chart represents a summary of all the meetings with a particular subject.

An example in Figure 4 presents an aggregated view on “hours of the day”. Each half-pie chart with colored stress levels represents a different hour of the day. It is important to notice, that this is not an hour of one day, but a summary across multiple days. The size of each hour represents the amount of SC measurements collected in this hours and each slice of the pie, represents a different stress level. It can be seen in the view, that the usual day for this user starts with low stress, between 7:00 and 8:00 and then the stress level increases till the maximum between 13:00 and 15:00. Another example of an aggregated view on “subjects” is presented in Figure 5, where each single half-pie chart represents an aggregation of stress levels during meetings with the same subject. It can be spotted there, that the “management” meeting takes a lot of time for a person and is more stressful than “lecture classes”. Furthermore, since people are better at reasoning based on instances rather than generalizations, we provide an interactive link between the aggregated views and the calendar such that when user click an aggregation “management”, the calendar view will display all the individual meetings with subject “management”. By providing such link, we enable people to better understand, trust and act on long term conclusions based on understanding of their physiological reactions in specific contexts.

Aside from the described views, the LifelogExplorer tool offers a number of additional functionalities that, while not necessarily for a regular user, become useful for an expert. One of such functionalities is the view on the histogram of collected data and the personalized stress level estimations. In this view, an expert can change the automated assignment of colors to SC measurements and define stress levels in a customized way.

Evaluation

We have conducted a field study with 21 vocational school teachers that were asked to use our tool for a period of 4 weeks and wear the DTI-2 measurement wristband during their work hours. They were able to visualize their data in the tool by connecting the device to the USB port of their computer at any convenient moment. After and during the study we collected their feedback about the use of the tool and their overall experience through questionnaires and interviews. As a result we have collected 20 days of measurements per participant on average ($\sigma = 5.92$) along with an average of 90 calendar entries per participant ($\sigma = 31.30$). We present only the results that directly relate to users’ use and experience with the tool. In total 5 of our users voluntarily asked for extending the study by 1 week. 2 of them indicated that they were interested to see if the patterns that they observed will repeat. Others did not provide a specific reason. 7 of the participants used the system on weekends. Those that indicated specific reasons, said that they wanted to see how doing the work related tasks on weekends affects them, or that they wanted to monitor a specific activity. Finally we asked for comments related to their experience with the tool. When asked what they liked about the system most indicated that it is easy and quick to understand and that there is not much you have to do. Also that it is a simple mechanism for learning your day routines quickly, and finally that the visualizations offer a good starting point for conversation. When asked what they did not like about it, most mentioned cumbersome measurement download from the device to the tool and that they kept forgetting to wear the measurement device and to turn it on. These comments suggest that the participants can find their own specific use of the tool and they can see the value it can give them, but still suggests that the measurement procedure could be improved. Voluntary use of the system, outside of requested times, suggests that the participants found the tool useful in practice.

Conclusions & Future work

We presented a tool for effective processing, interpretation and visualization of the long-term SC measurements with real-life context that is suitable for a non-expert user. The proposed visualization supports users in building up self-awareness about their stress responses in various contexts. In our future work we would like to investigate extending our tool with additional signal sources and more information about context.

Ethical statement & Acknowledgements

The studies performed as a part of this research project have been approved by the ethical committee of Philips Research - ICBE (Internal Committee Biomedical Experiments). We would like to thank all the school teachers for participating in our study and providing their comments.

References

[1] Bacharach, S.B., Bamberger, P., Conley, S. (1991). Work-home conflict among nurses and engineers: Mediating the impact of role stress on burnout. *Journal of Organizational Behavior,* ***12***, 39–53.

[2] Sauter, S.L., Murphy, L.R. (1995). *Organizational risk factors*. Washington, DC, USA: American Psychological Association.

[3] Kocielnik, R., Maggi, F., Sidorova, N. (2013). Enabling self-reflection with LifelogExplorer: Generating simple views from complex data. In *7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, Venice, Italy.

[4] Westerink, J., Ouwerkerk, M., de Vries, G., de Waele, S., van den Eerenbeemd, J., van Boven, M. (2009). Emotion measurement platform for daily life situations. In *Proceedings of the IEEE 3rd International Conference on Affective Computing and Intelligent Interaction, ACII*, Vol 1, p. 217-222. 10--12 September, Amsterdam, The Netherlands. IEEE Press.

[5] Kocielnik, R., Sidorova, N., Maggi, F., Ouwerkerk, M., Westerink, J. (2013). Smart technologies for long-term stress monitoring at work. In *IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS)*.

[6] Singer, E. (2011). The measured life. *Technology Review,* ***114(4)***, 38–45.

Watch it! A study into the associations between skin conductance and aggressive behavior

H. Nijman[1], P. de Looff[1], E. Kuijpers[2], J. à Campo[3], M. Stevens[2], M. Budy[3], P. Ligtvoet[1] and M. Ouwerkerk[4]

[1]Altrecht Aventurijn, Den Dolder, the Netherlands
[1] H.nijman@altrecht.nl
[2] GGZE, Eindhoven, the Netherlands
[2] erik.kuijpers@recornect.com
[3] Mondriaan GGZ, Maastricht, the Netherlands
[3] j.a.campo@mondriaan.eu
[4] Philips Research, Eindhoven, the Netherlands
[4] martin.ouwerkerk@philips.com

Introduction

Measurements of Skin Conductance Levels (SCLs), and short term increases of skin conductance in response to triggers (Skin Conductance Responses; SCRs), have been used to assess (psychological) arousal for well over a century now [1]. The number of studies of SCLs and SCRs of hospitalised psychiatric patients residing at (locked) psychiatric admission wards, however, is limited [but for instance see 2]. One of the reasons for this probably is that measuring skin conductance, until recently, required a substantial amount of equipment to which the participant had to be attached to obtain these assessments. Nowadays it is possible to measure (changes in) SCLs and SCRs by means of small wearable devices in the form of regular watches or wrist bands. Such devices can be used to assess various psycho physiological parameters, among which the SCL, without the patient being restricted (i.e., having to sit in one place). This provides more opportunities to measure psycho physiological changes in psychiatric patients during their stay at a psychiatric ward, even when they are in a state of acute psychiatric crisis [3]. By means of such a wearable device (see Photo 1), the associations between (changes in) the SCL and SCRs on the one hand, and agitated and aggressive behavior on the other, were studied among a group of acutely admitted (forensic) psychiatric patients. Aggressive behavior is a prevalent phenomena on (locked) psychiatric wards which threatens the safety and well-being of staff members and patients [for a review on the prevalence of aggressive behavior on psychiatric wards see 4].

Methods

In the current study, the arousal levels of 47 (forensic) psychiatric patients were preferably assessed during two full days (from 9 o clock AM till 5 o clock PM), or at least during one full day (i.e., for 15 of the 47 included patients only one day of assessments could be obtained). The patients were admitted either to the intensive psychiatric care ward of GGZE in Eindhoven, the Netherlands, the acute admissions ward of Mondriaan Mental Health institute in Maastricht, the Netherlands, or to the forensic psychiatric admissions ward Roosenburg in Den Dolder, the Netherlands. The SCLs of the participating patients were recorded at two Hertz by means of the Digital Tension Indicator – 2 (DTI-2; see Figure 1).

Each half hour, potential aggressive behavior was assessed by an observing staff member by means of the Social Dysfynction and Aggression Scale (SDAS) [5]. The SDAS consists of 11 items on which an observing staff member can document any agitated or verbal and physical aggressive behavior, ranging from mild to severe.

Statistical analyses

It concerns a within subject design in which the SCLs and SCRs during half hours in which aggression was observed were compared with the SCLs and SCRs during the half hour preceding the observation of aggression, as well as with the two consecutive half hours preceding the observation of aggressive behavior. A within subject was chosen as individuals have substantially different baseline SCLs in rest. Non parametric tests were used (i.e., the Wilcoxon Signed ranks test and Friedman's Anova) for the comparisons, as the SCLs were not normally distributed. For these comparisons, alpha was set at 0.05, two-tailed.

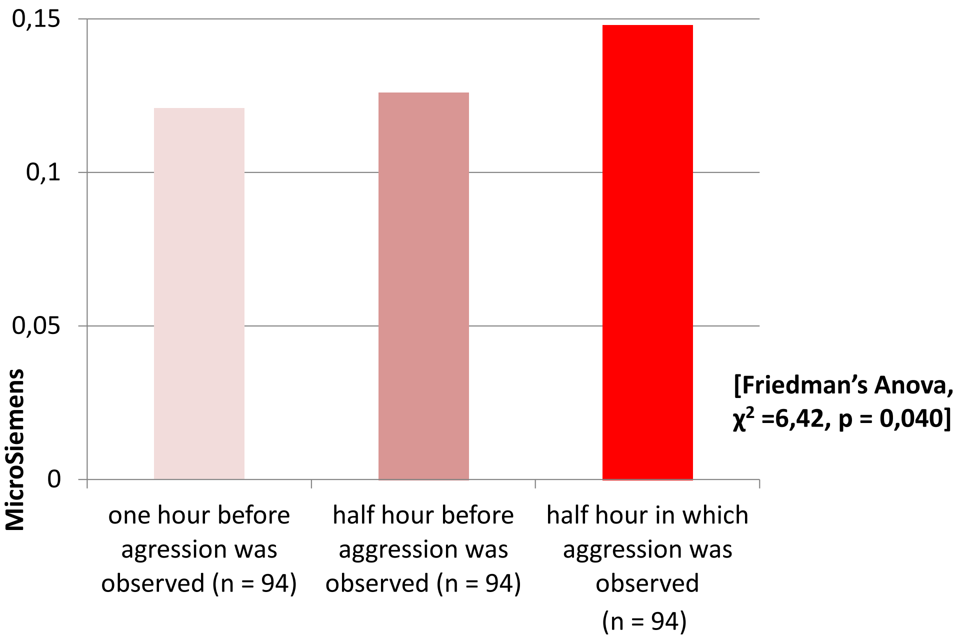
Photo 1. The Digital Tension Indicator – 2 (DTI-2) used in the current study to assess the skin conductance level of the participating patients.

Results

Of the 47 participants, 42 (89%) had exhibited aggressive behavior as assessed with the SDAS during at least one half hour period. This sample of 42 patients consisted of 36 men (86%) and 6 women (14%) The most prevalent psychiatric diagnoses in this sample of 42 patients were psychotic disorders (59%), cluster B-personality disorders (28%) and mood disorders (13%).

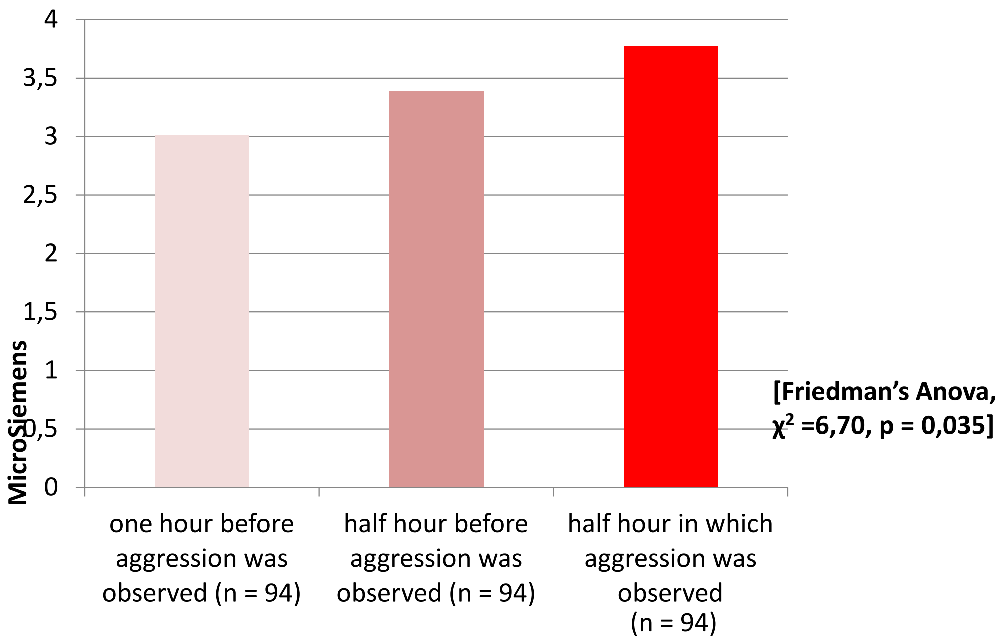
In Figure 1, the mean SCL during half hours in which the patients had displayed aggressive behavior is contrasted with the half hours preceding the observation of aggressive behavior.

Figure 1. The mean SCL during half hours in which aggression was observed(red bar) compared to the mean SCL in the preceding two half hour periods



In Figure 2, the number of SCRs during half hours in which the patients had displayed aggressive behavior is contrasted with the two half hours preceding the observation of aggressive behavior.

Figure 2. Mean number of SCRs (per minute) during half hours in which aggression was observed (red bar) compared to the number of SCRS in the preceding two half hour periods



Discussion

During half hours in which aggressive behavior had been observed, both the SCL and number of SCRs were higher (on average) when compared to the timeframe preceding these half hours. The differences, however, were limited in absolute terms, and the mean SCL of the patients as measured on the wrists of the participating patients with the DTI-2 in a general sense was low. Possibly, the measurement units for documenting aggression used in the current study, in which aggressive behavior was documented per half hour, are too crude to gain detailed insight in which potential changes in skin conductance are associated with aggressive behavior. A more precise documentation of the exact time the aggressive conduct started seems advisable for future studies on the associations between changes in electro dermal activity and aggressive behavior.

References

[1] Boucsein, W. (2012). *Electrodermal activity (2nd ed.).* New York: Springer.

[2] Campo, J.M.L.G., à, Merckelbach, H.L.G.J., Nijman, H.L.I., Allertz, W.W.F., Yeates-Frederikx, M. (2000). Schizofrene symptomatologie en huidgeleidingsactiviteit. *Acta Neuropsychiatrica*, ***12***, 177 - 182.

[3] Kuijpers, E. , Nijman, H., Bongers, I.M.B., Lubberding, M., Ouwerkerk, M. (2012). Can mobile skin conductance assessments be helpful in signalling imminent inpatient aggression? *Acta Neuropsychiatrica, **24***, 56–59.

[4] Nijman, H., Palmstierna, T., Almvik, R., Stolker, J. (2005). Fifteen years of research with the Staff Observation Aggression Scale; A review. *Acta Psychiatrica Scandinavica*, ***111***, 12 -21

[5] Wistedt B., Rasmussen A., Pedersen L. et al. (1990). The development of an observer-scale for measuring social dysfunction and aggression. *Pharmacopsychiatry*, ***23***, 249-252.

MoodRadar-DAVID: Facilitating insightful care by combining caretaker expertise and real time insight in electrodermal fluctuations

M.L. Noordzij[1] and M.E. Laroy-Noordzij[2]

[1]Department of Cognitive Psychology and Ergonomics, University of Twente, Enschede, The Netherlands
[1] m.l.noordzij@utwente.nl
[2]Dienst Behandeling en Zorg, De Twentse Zorgcentra, Enschede, the Netherlands.
[2] marleen.laroy@detwentsezorgcentra.nl

Introduction

Caretakers of people with severe mental handicaps report that the subjective interpretation of client behavior is difficult: Physical characteristics and behavioral signals are very heterogeneous. Identical client behavior is interpreted differently by different caretakers. Due to high workload (physically demanding chores for often more than one client at a time) the non-verbal signs of rising arousal levels are easy to overlook in real time situations. These behavioral signs are being labeled as important markers for the arousal build up leading up to outburst of challenging behaviors (e.g. aggressive acts) in retrospective video analysis.

Caretakers do indeed also explicitly report that they are often taken by surprise when a client shows challenging behavior. They express a need for additional tools to reliably and continuously track arousal states of their clients. Moreover, the clients who show unpredictable, challenging behavior often suffer dire consequences. For example, they get more medication, and they can be confronted with measures that constrain their freedom.

MoodRadar-DAVID

The MoodRadar-DAVID (MRD) project was set up to directly address the needs of the caretakers and increase well-being of clients. We aim to do this by creating a system that detects the continuous fluctuations of arousal by measuring electrodermal activity (EDA) of individual people with mental handicaps in their day-to-day life and that can alert caretakers when arousal is high (for that person). In addition, these physiological signals need to be aggregated and visualized in such a way that it provides the caretakers with relevant insight into the arousal state of their clients. Measures of EDA can be seen as a very pure measure of the activity of the sympathetic part of the autonomic nervous system, and as such can be taken as indicative of arousal levels of individuals [1]. In this project we focus on measuring skin conductance and extracting the most frequently reported parameters from this signal in ambulatory studies (such as the number of skin conductance responses per minute and the total amplitude of skin conductance response in a given minute)[1].

At this point it is important to note that the MRD project does not aim to use physiological signals to precisely classify complex emotional states or uniquely predict certain specific events in the future (such as an aggressive act). This would mimic mistakes made in other applied physiological domains such as some fields of lie detection (for a review see [2]). Here one could assume that, for example, a rise in skin conductance levels when answering a question indicates that someone is lying. This might very well be true, but the mistake is that this rise in skin conductance levels can also be expected when someone is afraid that their truthful answer might not be believed (for a review see [3]). Given our current scientific understanding of the interactions (or lack there-of) between human experience, observable behavior, cognitive functions, environmental conditions and (neuro)physiological states one cannot take fluctuations or regularities in (neuro)fysiological signals and assume that they will correspond uniquely with specific behavior or emotional experiences (for a very clear and still relevant introduction in associated cognitivist and phenomenological traditions see [4]).

The strength of our approach is that we take a demanding, ambiguous vigilance task (i.e. monitoring the arousal states of other persons on the basis of non-verbal signs) and complement this with monitoring technology, which aims to do the same task on the basis of a different source (changes in skin conductivity). The basic assumption here is that high arousal is one of the factors that is very often involved in the occurrence of and build-up towards challenging behavior. By providing caretakers with an accurate insight in the arousal of clients and by capturing their attention when arousal states move from low-medium to medium-high values (for an individual) MRD aims to alleviate some of the burden of caretakers. It does this by supporting them in their monitoring task and making sure their trained eye and caring expertise is, at minimum, focused on the client at relevant moments (i.e. when their clients' arousal is above average levels).

Implementation

From the start MRD was envisioned as a multi-stage design and research project where later stages were and are dependent on the success of earlier stages. Initially, we needed to establish whether an off-the-shelf ambulatory skin conductance sensor would suffice to measure skin conductance of the clients reliably during their normal lives, and whether relevant parameters (such as the rapid increases in the skin conductance levels denoted as skin conductance responses [1]) could be extracted from the measurements. In addition, we needed to find out whether the introduction of the wearable technology would not distress the clients even more, making the project immediately counter-productive relative to its aims. Previously we reported on the success of this first stage [5], and from there we started working on a suitable algorithm (programmed in Python) for aggregating and visualizing the skin conductance data (this will be reported elsewhere).

At this point we also started to focus on an initial prototype of the MRD system. For this we relied on practices from User-Centered Design involving important stakeholders to arrive at (a prioritization of) requirements for the system and a persona (a fictional description of characteristics and tasks of a relevant end user, which can be very helpful for software developers) [6]. The algorithm, requirements and the persona were taken as input for the construction of a MRD app through an agile development cycle (of approximately 12 weeks) lead by the company Move4Mobile (http://www.move4mobile.com/nl/).

Evaluation

At present a 3 month evaluation period is taking place of the MRD app (up to date information on the progress of the project will be provided at http://buienradar.detwentsezorgcentra.nl). Ethnographic methods and surveys will be used to assess the actual use, usability and user satisfaction with approximately 30 caretakers. In addition, three clients will participate in three separate single-case ABA research design [7] to establish whether an intervention with MRD has positive effects on their well-being. Firstly, announcing client arousal changes will hopefully allow a window of opportunity to arise to adjust caretaking attitude in order to prepare for, or to prevent challenging behavior from escalating, or occurring at all, through existing intervention techniques. Secondly, MRD aims to enhance the feelings of predictability in the caretaker in relation to the occurrence of challenging client behaviors and thereby diminish the feelings of caretaker stress and make negative, freedom constraining actions towards the client unnecessary.

Ethical Statement

This study was approved by the local Medical Ethical Committee, MST hospital, Enschede (METC no. P11-27 NL 37314.044.11, approved on 06-09-2011 & METC no. P13-32 NL 46166.044.13, approved on 18-02-2014). This study was also registered in the Netherlands Trial Register (Trial Code 3043).

Acknowledgements

This study has been supported by "De Twentse Zorgcentra", and by Zorgkantoor/Menzis regio Enschede.

References

[1] Boucsein, W. (2012). *Electrodermal Activity (2nd ed.).* New York, NY, USA: Springer.

[2] Vrij, A. (2008). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Wiley-Interscience.

[3] Iacono, W.G. (2007). The *Detection of Deception*. Cambridge University Press, City.

[4] Varela, F., Thompson, E., Rosch, E. (1992). *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press.

[5] Noordzij, M.L., Scholten, P., Laroy-Noordzij, M.E. (2012). Measuring electrodermal activity of both individuals with severe mental disabilities and their caretakers during episodes of challenging behavior. In *Proceedings of the Measuring Behavior 2012* (Utrecht, The Netherlands).

[6] Hein, I.C.A. (2013). *Observing arousal at mentally challenged clients in a health care environment: a user requirements study for a monitoring device that support care takers with working with their clients.* University of Twente, Enschede, The Netherlands.

[7] Dugard, P., File, P., Todman, J. (2012). *Single case and small-N experimental designs. A practical guide to randomization tests.* New York, NY, USA: Routeledge.

Measuring Depression-Relevant Behavioural States in Mouse Models for the Inflammation Hypothesis of Depression

C.R. Pryce[1], D. Azzinnari[1], G. Bergamini[1], F. Klaus[1], F. Cathomas[1], R. Fuertig[2], A. Ceci[2], B. Hengerer[2]

[1] PLaTRAD, Department of Psychiatry, Psychotherapy & Psychosoamtic, University of Zurich, Zurich, Switzerland

[1] christopher.pryce@bli.uzh.ch

[2] CNS Disease Research, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany

Behavioural states valid for depression

Despite their oft-stated importance to understanding human mental illness and developing novel drug therapies, in practise, animal models of psychiatric disorders frequently lack relevance (validity) in terms of aetiology and/or psychopathology. In depression, for the major domains of stress-induced generalised helplessness and fatigue, there are no animal models. That rodent emotional-cognitive function is sufficiently complex to justify the study of depression-relevant processes is exemplified by their sensitivity to stimulus un/controllability: mice that experience a stressor as inescapable develop control deficits (i.e. fail to escape) when it is escapable [1]. However, such specific learned helplessness is of limited relevance to human helplessness where it is the inability to control one stressor, often social, leading to generalised helplessness that is important [2]. Generalised helplessness constitutes changes in three underlying emotional-cognitive states: increased emotional reactivity to aversive stimuli e.g. increased freezing to stimuli conditioned to electroshock; decreased motivation to respond actively to aversive stimuli e.g. decreased reactivity to electroshock; decreased cognitive operant learning (and expectancy) that behavioural responses terminate aversive stimuli. Using a purpose-built multi-conditioning system [1] it is possible to measure each of these three behavioural states with one system. Although fatigue is a core depression symptom, there is no animal model for stress-related fatigue. Using a treadmill combined with electroshock, fatigue induced by enforced running can be studied in mice. This has been applied in inflammation and exercise physiology and we have used this readout to study psychosocial stress-induced fatigue. Reduced reward sensitivity is typically studied using the two-bottle sucrose/saccharin vs. water preference test; the test is often conducted after a period of enforced water deprivation which will yield false positive effects if the manipulation induces increased allostatic fluid needs as occurs, for example, with chronic social stress. For accurate measurement of reductions in reward motivation, as well as changes in emotional-cognitive processing of reward, operant paradigms are necessary [3, 4]. Such operant tests can be conducted either in remote testing chambers or in the home environment using IntelliCage.

Chronic social defeat induces inflammation and depression-relevant behavioural states

Adult male mice were exposed to 15-day uncontrollable psychosocial stress without physical wounding, i.e. solely emotional stress, and investigated in terms of behavioural, physiological and neurobiological effects [5]. Relative to controls, socially-stressed mice displayed altered reactivity to a physical stressor (mild electroshock) in terms of increased fear acquisition, generalised helplessness, and increased fatigue. In terms of reward-based behaviour, they displayed decreased operant responding for saccharin solution and water, no change in saccharin drinking but increased water drinking. They also displayed decreased motivation for sucrose pellet reward on a progressive ratio schedule of reinforcement, and increased sensitivity to learned non-reward and impaired probabilistic reversal learning in operant tests with sucrose-pellet reward. Alongside these depression-relevant behavioural states, socially-stressed mice exhibited increased blood levels of pro-inflammatory cytokines and spleen hypertrophy accompanied by increased splenic granulocytes, inflammatory monocytes and T helper 17 cells, and increased blood and brain levels of tryptophan catabolites e.g. kynurenine. Central nervous system gene expression was altered in pathways of inflammation, G-protein coupled receptors and dopamine function. The increased conditioned fear expression induced by chronic social defeat was reversed by 3-day administration of an inhibitor of indoleamine 2,3-dioxygenase.

Peripheral tumor necrosis factor induces inflammation and depression-relevant behavioural states

An adeno-associated (AAV) murine TNF vector was produced to investigate effects of CNS region-specific TNF over-expression on depression-relevant phenotypes. At 14-days post-stereotactic injection of 1.1e9-1.1e12 vg/ml TNF vector in hippocampus, there was a localized, dose-dependent increase in immunoactive TNF. The titre 1.1e10 vg/ml yielded 0.5-100 pg/mg protein (vs. 0.3 pg/mg in control mice), was without effect on brain cytoarchitecture or peripheral TNF levels, and was selected for study. Effects of bilateral 14-day AAV TNF over-expression in the amygdala were studied in the fear conditioning paradigm. There were no effects on either acquisition or expression of conditioned fear. In contrast, 6-day peripheral TNF administration increased acquisition and expression of conditioned fear. Effects of a CD40 agonist antibody were investigated in terms of behavioural, physiological and neurobiological effects. Acute effects were increased plasma TNF levels and anorexia. Persistent physiological effects included increased blood and brain levels of tryptophan catabolites e.g. kynurenine. Persistent behavioural effects included decreased saccharin consumption in the two-bottle test and decreased operant responding for saccharin solution and water. These anhedonia effects of CD40-TNF activation could be prevented by the co-administration of the TNF inhibitor etanercept. In contrast to peripheral TNF, CD40-TNF activation induced impaired classical conditioning of fear, suggesting that the increased fear reactivity induced by TNF per se can be masked by immune-inflammatory impairment of synaptic plasticity.

References

[1] Pryce, C.R., Azzinnari, D., Sigrist, H., Gschwind, T., Lesch, K.-P., Seifritz, E. (2012). Establishing a learned helplessness effect paradigm in C57BL/6 mice: behavioural evidence for emotional, motivational and cognitive effects of aversive uncontrollability per se. *Neuropharmacology, **62***, 358-372.

[2] Pryce, C.R., Azzinnari, D., Spinelli, S., Seifritz, E., Tegethoff, M., Meinlschmidt, G. (2011). Helplessness: a systematic translational review of theory and evidence for its relevance to understanding and treating depression. *Pharmacology & Therapeutics, **132***, 242-267.

[3] Ineichen, C., Sigrist, H., Spinelli, S., Lesch, K.-P., Sautter, E., Seifritz, E., Pryce, C.R. (2012). Establishing a probabilistic reversal learning test in mice: evidence for the processes mediating reward-stay and punishment-shift behaviour and for their modulation by serotonin. *Neuropharmacology, **63***, 1012-1021.

[4] Spinelli, S., Müller, T., Friedel, M., Sigrist, H., Lesch, K.-P., Henkelman, M., Rudin, M., Seifritz, E., Pryce, C.R. (2013). Effects of repeated adolescent stress and serotonin transporter gene partial knockout in mice on behaviors and brain structures relevant to major depression. *Frontiers in Behavioral Neuroscience, **7**(215)*, 1-15.

[5] Azzinnari, D., Sigrist, H., Staehli, S., Palme, R., Hildebrandt, T., Leparc, G., Hengerer, B., Seifritz, E. (2014). Mouse social stress induces increased fear conditioning, helplessness and fatigue to physical challenge together with markers of altered immune and dopamine function. Neuropharmacology, *in press*.

The role of Neutrophil Gelatinase Associated Lipocalin in linking Depression and Heart Failure.

Leonie Gouweleeuw[1], Uli LM Eisel[1,2], Pieter JW Naude[1], Regien G. Schoemaker[1,3]

[1] Department of Molecular Neurobiology, University of Groningen, Groningen, the Netherlands
[2] Department of Psychiatry, University Medical Center Groningen, Groningen, the Netherlands
[3] Department of Cardiology, University Medical Center Groningen, Groningen, the Netherlands

Introduction

Depression is a well-known co-morbidity in heart failure. In fact, approximately half of all patients with heart failure suffer from signs of depression. Co-morbid depression severely worsens prognosis in these patients. However, the majority of these patients remain untreated; beside the point that the cardiologist regards depression often as a “natural” response on a life threatening condition, a rational for adequate therapy is hampered by a lack of knowledge concerning the underlying mechanism of this neurocardiac interaction. Moreover, optimal cardiovascular treatment is not associated with lower incidence of depression, while the other way around, successful anti-depressant therapy is not associated with improved prognosis. These observations may suggest a common mechanism, rather than a direct causal relationship underlying heart failure and depression.

Inflammation may be an underlying biological mechanism and a common denominator in heart failure as well as depression. Heart failure and depression share increased circulating levels of pro-inflammatory cytokines, including TNFα, while co-morbidity is associated with higher levels than each condition alone. We hypothesize that an initial functional immune response on local tissue damage, is not able to keep this process local, either due to repeated exposure or due to an exaggerated inflammatory response (reviewed by Quan, 2014(1)). Hence, immunosensitive afferent nerves can be stimulated, as well as inflammatory factors may enter the circulation, both of which may enter the brain either directly or via the blood-brain-barrier, and subsequently induce neuroinflammation. Neuroinflammation could provide protection to limit CNS damage, but may confer long-term consequences related to chronic exposure to inflammation associated neurodegeneration. The localization of these processes in the brain then determines the specific behavioral output. This hypothesis is presented for myocardial infarction in figure 1.

Studies in rats with myocardial infarction support a role for TNFα in neuroinflammation with depression and heart failure. Myocardial infarction evokes elevated levels of cardiac as well as circulating TNFα, starting 30 minutes after MI, and substantially further increases the following weeks. We previously showed that myocardial infarction induces focal leakage of the blood-brain-barrier at least up to 3 days post-infarct; an effect that can be mimicked by peripheral TNFα infusion. In the brain, amongst others, TNFα expression is increased at mRNA level as well as protein level in the hypothalamus, but not the cortex. In the paraventricular nucleus of the hypothalamus, microglia activation is observed up to 16 weeks after myocardial infarction. Moreover, in these rats, signs of depressive behavior are reported, that can be inhibited by TNFα blockade (etanercept) (2). However, while the hypothalamus can be regarded as a relative non-specific brain area in relation to more complex behavioral changes as depression, TNFα can be regarded as an important, though relative non-specific cytokine, as it is activated in numerous conditions. Therefore the aim of our research is to further unravel the underlying mechanisms involved in the heart failure-(neuro)inflammation-depression triangle.

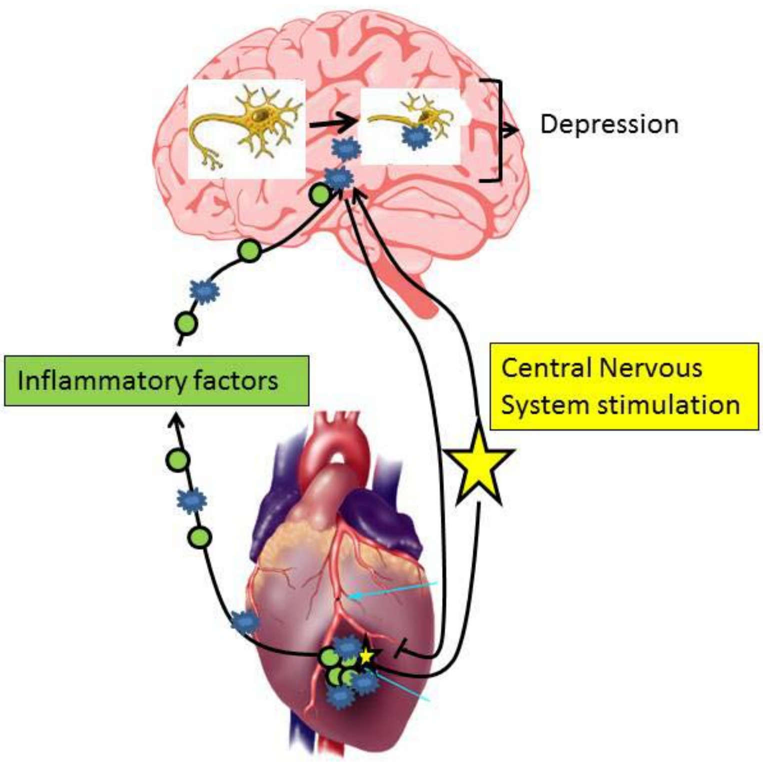


Figure 1. Hypothesis concerning the role of inflammation in the development of heart failure and depression. A local inflammatory response on tissue damage due to myocardial infarction evokes activation of inflammatory cells and factors to "heal" the infarct. This initially local inflammatory response can be expanded into the brain either by stimulation of afferent nerves, or by entering the circulation and crossing the blood brain barrier. Localized inflammatory responses in the brain may then be associated with depressive behavior.

A detailed experiment investigating novel targets induced by TNFα, revealed Neutrophil Gelatinase Associated Lipocalin (NGAL) as an interesting candidate. It can be produced by neurons, microglia and astrocytes, and is activated throughout the brain at peripherally induced inflammation by LPS. Moreover, as NGAL is a well-recognized predictive factor for prognosis in heart failure, it is also found increased in late life depression (3). Most recently we showed a significant association of depressive symptoms in patients with heart failure and increased plasma NGAL levels, even after correction for cardiac- and renal dysfunction(4). This association, however, was observed for the somatic, rather than the cognitive aspects of depression, indicating a specific subset of elements of depression to be affected.

In the present study we investigated 3 elements of depression after myocardial infarction in rats, and evaluated a potential role of NGAL.

Methods

Experimental myocardial infarction in rats.

Male Wistar rats were subjected to experimental myocardial infarction by surgical ligation of the left descending coronary artery or sham surgery. Ligation results in a massive transmural infarct in the free wall of the left ventricle leaving papillary muscles largely unaffected, hence leading to pump failure rather than cardiac arrhythmias.

Behavioral tests

Previously, we showed that rats with myocardial infarction showed depressive-like behavior, presented as decreased interest in their environment, measured as lower time spent on exploratory behavior in an open field test, and lower social interest, measured as time spent on social interaction (5). Therefore, 3 weeks after myocardial infarction or sham surgery, when irreversibly damaged cardiac tissue is replaced by scar tissue, behavioral testing was performed. “Depressive-like” behavior was evaluated from exploration in the open field, as measures for interest in environment; latency to leave the home cage to enter a novel area, as measure of motivation; and a sucrose preference test, as measure for loss of pleasure; anhedonia. All behavioral tests were performed during the dark phase; the active period of the rat.

For the open field test: rats were placed in the center of the open field (1m x1m) and were allowed to explore it for 5 minutes. Behavior was recorded and a total review of the rats behavior was obtained off-line using E-line software. Time spent on the different behavioral elements; sniffing, walking, rearing, exploring wall, resting, grooming, other, was calculated and time spent on exploration was regarded the summation of the first 4 elements. Less time spent on exploration is regarded as lower interest in environment; a sign of depression.

Free exploration: in this test we gave the rat a choice to whether or not leave its home cage and explore the open field. For that the home cage with the rat in it was placed in the open field, the lit was removed, and the rats was allowed for 5 minutes to climb out and explore the open field. The latency to leave the home cage was recorded and a higher latency is regarded as sign of depression; a lower motivation to explore a new environment (5).

Anhedonia was obtained from a sucrose preference test. For that, after a habituation session of 2 bottles and the taste of sucrose, rats were subjected to the choice of normal water or 1% sucrose for the first 5 hours of the dark period (period in which rats drink most). A lower percentage of sucrose of the water intake is regarded as sign of depression.

Although the used behavioral tests, except for the free exploration, may be used rather generally, this combined setting may provide a novel and relevant way to study the mechanism underlying “depression”, which clinically is a highly variable combination of symptoms.

Tissue and plasma analyses

Five weeks after myocardial infarction or sham surgery, rats were sacrificed. Plasma and cerebrospinal fluid samples (CSF) were collected for measurement of NGAL levels. Subsequently the brain was perfused with formaldehyde for fixation and processed for immunohistochemical staining of NGAL. The heart was dissected and processed for measurement of infarct size as the % of left ventricular circumference at mid-ventricular level.

Experimental procedures were approved by the local animal experiments and welfare committee of the University of Groningen, the Netherlands.

Results

In the present study, rats with myocardial infarction did not show statistically significant less exploration in the open field test. Latency to leave the home cage appeared higher in rats with myocardial infarction (74±12 versus 58±9 sec). Difference in sucrose preference, 77±6% in infarcted versus 83±4% in sham rats, did however not reach statistical significance. Time spent on exploration in the open field was not correlated to either latency to leave the home cage or sucrose preference. Sucrose preference and latency to leave the home cage were significantly correlated ($p=0.024$). However, this correlation was found positive, leaving the rats with highest sucrose preference displaying the longest latency.

Plasma levels, as well as CSF levels of NGAL, were higher in infarcted compared to sham rats (plasma: 170±37 versus 115± 25 ng/ml; CSF 25.9±17.2 versus 6.5±2.9 ng/ml), but did not reach statistical significance. The number of NGAL positive cells in the magnocellular part of the paraventricular nucleus of the hypothalamus (PVN), but not in the parvocellular part, was significantly increased after myocardial infarction (45±7 versus 27±2). No differences were observed in the different areas of the hippocampus, nor in the prefrontal cortex.

Interestingly, whereas plasma levels of NGAL were significantly correlated to infarct size, but not to brain NGAL expression, NGAL expression in the magnocellular part of the PVN appeared significantly correlated to exploration in the open field, while correlations with latency (p=0.100) and sucrose preference (p=0.053) did not reach statistical significance, but showed a tendency. Moreover, magnocellular expression of NGAL was significantly correlated to the number of microglia in the PVN (p=0.023).

Discussion/conclusion

In the present study, 3 different elements of depressive-like behavior were evaluated in a rat model for myocardial-infarction associated depression; loss of interest in environment, lack of motivation and anhedonia. Not all measured elements of depressive-like behavior did statistical significantly differ between sham and infarcted rats, though a clear tendency was present. Since within the same animals, scores of these elements were either not correlated, or correlation was even in the opposite direction, the elements may indeed reflect different (and none-related) symptoms of depression. Since different, and even opposing effects are also observed in patients with depression (DSM-V), combining different elements of depression using different (non-aversive) tests in the same animal, may provide a more relevant approach to investigate the complex syndrome of depression in animal models, than the common single-test studies.

Experimental myocardial infarction in rats increases production of NGAL, resulting in increased plasma levels, related to infarct size. This indicates that plasma NGAL levels may reflect severity of cardiac dysfunction. However, since no correlation between plasma levels of NGAL with brain expression of NGAL was observed, nor with any measured aspect of depressive behavior, plasma NGAL may not directly affect brain functioning. Expression of NGAL in the brain was significantly increased in the PVN, which correlated to exploration behavior in the open field while a tendency towards correlation with the other depression parameters were observed. This may indicate that increased brain NGAL expression may reflect ”depressive like behavior” associated with heart failure rather than heart failure itself.

Both circulating and brain-associated NGAL may reflect different aspects of heart failure that are important for prognosis. The lack of consistent associations between circulating and brain NGAL may support the idea of a common underlying mechanism, rather than a causal relationship for heart failure and co-morbid depression. Since NGAL may play an important role in inflammatory processes, the study may provide evidence for our neuroinflammation hypothesis.

References

[1] Quan, N. (2014). In-Depth Conversation: Spectrum and Kinetics of Neuroimmune Afferent Pathways. *Brain, Behavior, and Immunity*. (*in press*)

[2] Grippo, A. J., J. Francis, R. M. Weiss, R. B. Felder, and A. K. Johnson. (2003). Cytokine Mediation of Experimental Heart Failure-Induced Anhedonia. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, ***284****(3)*, R666-73.

[3] Naude, P. J., U. L. Eisel, H. C. Comijs, N. A. Groenewold, P. P. De Deyn, F. J. Bosker, P. G. Luiten, J. A. den Boer, and R. C. Oude Voshaar. (2013). Neutrophil Gelatinase-Associated Lipocalin: A Novel Inflammatory Marker Associated with Late-Life Depression. *Journal of Psychosomatic Research*, ***75*** *(5)*, 444-450.

[4] Naude, P. J., P. M. Mommersteeg, W. P. Zijlstra, L. Gouweleeuw, N. Kupper, U. L. Eisel, W. J. Kop, and R. G. Schoemaker. (2014). Neutrophil Gelatinase-Associated Lipocalin and Depression in Patients with Chronic Heart Failure. *Brain, Behavior, and Immunity*. (*in press*)

[5] Schoemaker, R. G. and J. F. Smits. (1994). Behavioral Changes Following Chronic Myocardial Infarction in Rats. *Physiology & Behavior*, ***56***, 585-589.

Unraveling the Role of Neuroinflammation in Surgery-induced Cognitive Impairment

I.B. Hovens[1], R.G. Schoemaker[1,2], E.A. van der Zee [1] and B.L. van Leeuwen[3]

[1]Department of Molecular Neurobiology, University of Groningen, Groningen, the Netherlands
i.b.hovens@rug.nl
[2] Department of Cardiology, University Medical Center Groningen, Groningen, the Netherlands
[3] Department of Surgery and Surgical Oncology, University Medical Center Groningen, Groningen, the Netherlands

Abstract

Although inflammatory processes are generally considered to be involved in postoperative cognitive dysfunction (POCD), the mechanism behind this condition is still largely unknown. The aim of our research is to study the role of inflammation in the development of POCD under high and low risk conditions with special focus on the different brain areas and different types of cognitive function involved. To this end we developed a rat model for POCD using standardized surgery and various behavioral tests to assess surgery-induced cognitive dysfunction.

Introduction

Postoperative cognitive dysfunction (POCD) was first recognized as the occurrence of dementia-like symptoms that persisted for weeks to months after surgery [1]. Whereas in younger patients the cognitive impairments are usually transient, in approximately 10 % of the elderly patients POCD persists [2, 3]. Next to advanced age the main risk factors for POCD include surgery severity and duration, the occurrence of peri-operative complications and pre-existing cognitive impairment. POCD has been associated with reduced quality of life, increased disability and dependency on social services and an increased risk of mortality [4].

Consistent evidence is accumulating for the role of surgery-induced inflammation in the development of POCD. It is hypothesized that surgery leads to a systemic increase in pro-inflammatory cytokines, which in turn can influence inflammatory processes in the brain, resulting in microglia activation and neuroinflammation (see Figure 1). Indeed, several clinical and preclinical studies show a co-occurrence of post-operative (neuro)inflammation and cognitive impairment [5, 7–9]. Since this proposed mechanism is expected to occur to some extent in all patients, it remains unclear why some individuals develop persisting POCD and others do not.

Although much research has focused on determining the risk factors and mechanism for POCD, the cognitive domains affected by POCD are an understudied topic [10]. Clinical studies tend to define POCD as a persisting, generalized decline in cognition, without specifying which cognitive functions are impaired. Preclinical studies, mainly focus on surgery-induced hippocampal dysfunction. We reviewed the clinical literature on POCD with the aim to elucidate what cognitive domains deteriorate after surgery and what brain areas might be involved [10]. We concluded that POCD encompasses a wide range of cognitive impairments, suggesting POCD affects a multitude of brain areas. We therefore propose a new approach in (preclinical) research on POCD that encompasses various aspects of cognition and mood related behavior and their associated brain regions. We believe that this new approach will facilitated translation from studied in animal models to clinical practice.

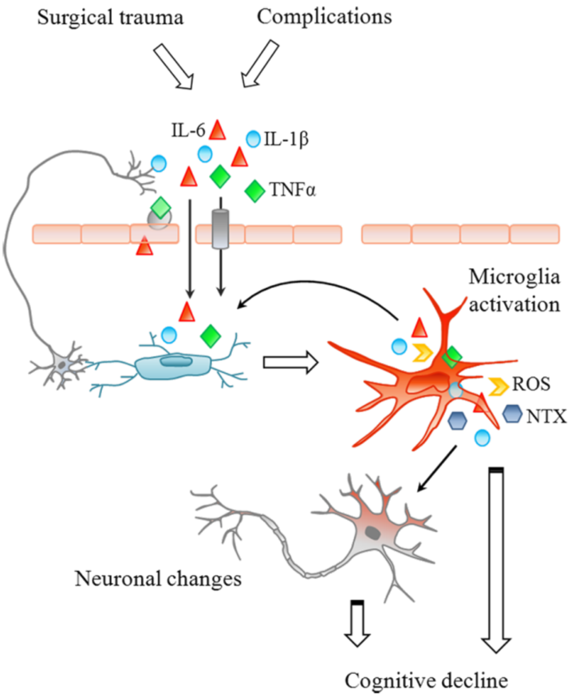


Figure 1. The hypothesized role of inflammation in POCD development. The surgical trauma and complications associated with surgery lead to a systemic increase in pro-inflammatory factors (e.g. IL-1β, IL-6 and TNFα). These can cross the blood brain barrier at permeable regions or by specific transport systems, or they can exert their effects in the CNS through interaction with vagal nerves. In the CNS inflammatory factors activate microglia, which in turn start producing more cytokines, reactive oxygen species and neurotoxins, ultimately leading to changes in neuronal functioning and cognitive decline (based on [5–7]).

Methods

We aimed to study the role of inflammation in the development of POCD under high and low risk conditions with special focus on the different brain areas and different types of cognitive function involved. To this end we developed an animal model for POCD using standardized abdominal surgery and various behavioral tests to assess surgery-induced cognitive dysfunction. We studied the effects of known risk factors for POCD such as time after surgery, increased age, pre-operative health status and the type of surgery on POCD, (neuro)inflammation and neuronal functioning. In this abstract we describe our outcomes of surgery on behavior and microglia activation.

Animals

To study development of POCD male Wistar rats of 3 months of age (young) were obtained from Harlan (Harlan laboratories, Venray, The Netherlands). To study the effects of aging male Wistar rats of 3, 18 and 25 months (young, aged and elderly respectively) of age were obtained from a breeding colony of the Semmelweis University (Budapest, Hungary). All experiments were approved by the local animal experiment and welfare committee (Dier Experimenten Commissie, Groningen, the Netherlands).

Surgery

Under sevoflurane-O_2 anesthesia and buprenorphine (0.003 mg/ kg s.c.) analgesia the gastrointestinal tract was exteriorized and a temporary hypoperfusion of the upper part of the mesenteric vascular bed was induced by clamping the upper mesenteric artery for 30 minutes [11]. In addition, the rats were equipped with an indwelling jugular vein catheter to allow timed blood sampling with minimal handling.

Behavioral tests

We performed several behavioral tests aimed at measuring different aspects of cognition and mood related behavior. All behavioral tests were performed following standardized protocols as described before [11, 12]. Open field (OF) testing was performed to analyze exploratory behavior in a novel environment as measure of interest and anxiety. The novel object (NO) test was used to determine attention and object memory. Spatial learning and memory were analyzed based on training and probe trials in the Morris water maze (MWM) or Y-maze. Finally, reversal training in the Morris water maze or Y-maze was used to determine cognitive flexibility.

Analysis of micorglia activation in IBA-1 stained brain sections

We developed a novel image-analysis image analysis method that uses morphological parameters of microglia to quantify microglia activation in IBA-1 stained brain sections. We studied the microglia activation in several brain regions that have been associated with aspects of mood and cognition. These areas are: the amygdala, which has been related to anxiety; the prefrontal cortex, which has been associated with object recognition; the hippocampus, which has been related to spatial learning and memory; and the striatum, which has been associated with cognitive flexibility.

Results and Discussion

Figure 2 gives a typical example of behavioral test outcomes and results for microglia activation obtained in our experiments.

<u>Young rats show impaired spatial memory and hippocampal neuroinflammation following surgery</u>

We studied the development of POCD in young healthy rats by analyzing behavior in the first three weeks following major abdominal surgery [11]. Anesthesia alone did not affect cognitive performance at any time point. In young rats spatial learning and memory were temporarily impaired in the first two weeks following surgery, whereas non-spatial cognitive functions seemed unaffected. Since spatial learning and memory depend mainly on hippocampal function, this finding suggests that the hippocampus may be especially vulnerable to surgery-induced impairment.

Microglia activation, indicative of neuroinflammation, was increased in the hippocampus and prefrontal cortex 1 week following surgery, but not thereafter. Comparing this to the behavioral test results, it seems that neuroinflammation is not the sole factor leading to POCD. We hypothesize that for POCD to occur neuroinflammation must affect pathways involved in neuronal functioning.

<u>Factors that are associated with an increased risk of POCD lead to more generalized and prolonged behavioral impairment following surgery</u>

Advanced age and cardiac surgery both increase the risk of POCD. In two separate experiments we compared POCD in young and elderly rats, and compared POCD in young rats after major abdominal and major cardiac surgery (unpublished results). Whereas young rats after abdominal surgery only showed impaired spatial memory, both elderly rats and rats that underwent cardiac surgery showed a more generalized cognitive decline including impairment of object recognition (NO) and cognitive flexibility (MWM). Future studies may indicate whether this more generalized cognitive impairment is related to a more generalized or prolonged neuroinflammation in the rat brain.

Whereas patients of all age categories can experience transient POCD, persisting POCD is mainly seen in elderly patients. We performed an additional experiment to study long-term postoperative behavioral changes in young and aged rats [12]. In contrast to young rats, aged rats showed a change in exploratory behavior that persists for at least 6 weeks following surgery. This may be indicative of changes in mood. Additionally, only aged rats showed increased microglia activation in the hippocampal CA1 region at six weeks following surgery, which was correlated to the changes in exploratory behavior. Again, these outcomes seem to indicate that the hippocampus is especially vulnerable to surgery-induced impairment. Moreover, a prolonged hippocampal neuroinflammation in aged rats may be associated with post-operative behavioral impairment.

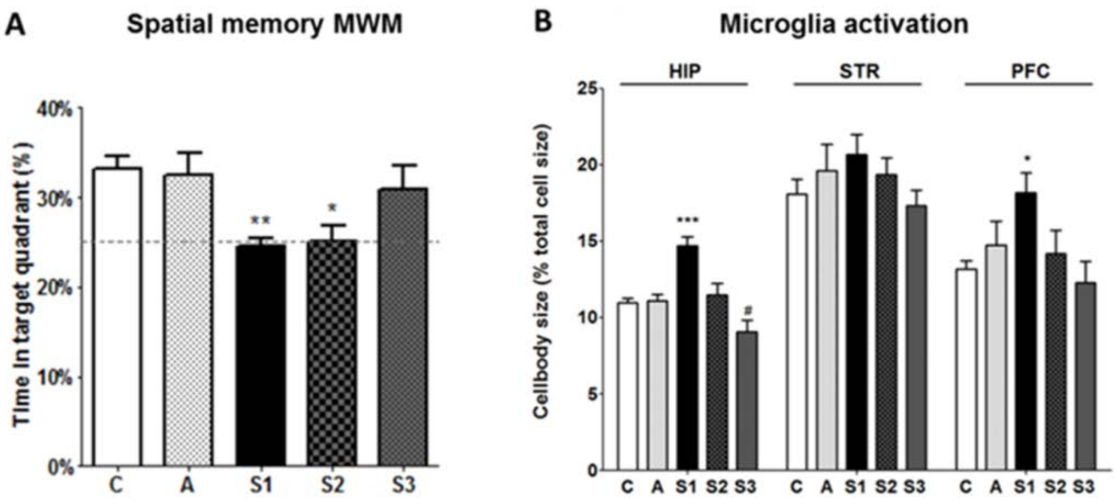


Figure 2. Spatial learning and microglia activation in young rats following abdominal surgery. A) Spatial memory in the Morris water maze (MWM) probe trial was measured by determining the time the rats spent in the target quadrant. Rats showed impaired spatial memory in the first and second week following surgery. Rats that only underwent anesthesia or rats that were tested in the third week following surgery performed similar to control rats. B) Microglia activation in the hippocampus (HIP), striatum (STR) and prefrontal cortex (PFC) was measured by determining the cell body to cell size ratio of IBA-1 stained microglia. Microglia activation was increased in the hippocampus and prefrontal cortex in the first week following surgery. C=naïve control group, A= anesthesia only group, S1, S2 and S3 = experimental groups that underwent behavioral testing 1, 2 and 3 weeks following surgery respectively. *p<0.05, **p<0.01,***p<0.001, #p<0.1 (based on [11]).

Conclusion

Our experiments are the first to study the effects of surgery on an extended range of cognitive functions and brain areas under high and low risk conditions for POCD. In line with previous research, we showed a temporal postoperative impairment of hippocampal dependent cognition and increase in hippocampal neuroinflammation in young healthy rats. Our approach allowed us to compare these outcomes in different cognitive domains. The results may indicate that in young rats hippocampal dependent cognition is most vulnerable to the effects of surgery and that neuroinflammation may not be the sole factor responsible for POCD. Additionally, we observed a more generalized and prolonged behavioral impairment and prolonged hippocampal neuroinflammation after surgery under conditions associated with increased POCD risk. These findings seem to be in line with the clinical findings on POCD and thus provide a promising basis for more extensive studies of the molecular changes in the brain following surgery.

References

[1] Bedford, P.D. (1955). Adverse cerebral effects of anaesthesia on old people. *Lancet*, ***269,*** 259–263.

[2] Rasmussen, L.S. (2006). Postoperative cognitive dysfunction: incidence and prevention. *Best Practice & Research: Clinical Anaesthesiology,* ***20***, 315–330.

[3] Moller, J.T., Cluitmans, P., Rasmussen, L.S., Houx, P., Rasmussen, H., Canet, J., Rabbitt, P., et al. (1998). Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International Study of Post-Operative Cognitive Dysfunction. *Lancet,* ***351***, 857–861.

[4] Steinmetz, J., Christensen, K.B. Lund, T. Lohse, N., Rasmussen, L.S., ISPOCD Group. (2009). Long-term consequences of postoperative cognitive dysfunction. *Anesthesiology*, ***110***, 548–555.

[5] Cibelli, M., Fidalgo, A.R., Terrando, N., Ma, D., Monaco, C., Feldmann, M., Takata, M., et al. (2010). Role of Interleukin-1Beta in postoperative cognitive dysfunction. *Annals of Neurology,* ***68***, 360–368.

[6] Krenk, L., Rasmussen, L.S., Kehlet, H. (2010). New insights into the pathophysiology of postoperative cognitive dysfunction. *Acta Anaesthesiologica Scandinavica,* ***54***, 951–956.

[7] Wan, Y., Xu, J., Ma, D., Zeng, Y., Cibelli, M., Maze, M. (2007). Postoperative impairment of cognitive function in rats: a possible role for cytokine-mediated inflammation in the hippocampus. *Anesthesiology,* ***106***, 436–443.

[8] Li, X., Wen, D.-X., Zhao, Y.-H., Hang, Y.-N., Mandell, M.S. (2013). Increase of beta-amyloid and C-reactive protein in liver transplant recipients with postoperative cognitive dysfunction. *Hepatobiliary & Pancreatic Diseases International,* ***12***, 370–376.

[9] Jones, E.L, Gauge, N., Nilsen, O.B., Lowery, D., Wesnes, K., Katsaiti, E., Arden, J., et al. (2012). Analysis of neuron-specific enolase and S100B as biomarkers of cognitive decline following surgery in older people. *Dementia and Geriatric Cognitive Disorders,* ***34***, 307–311.

[10] Hovens, I.B., Schoemaker, R.G., van der Zee, E.A., Heineman, E., Izaks, G.J., van Leeuwen, B.L. (2012). Thinking through postoperative cognitive dysfunction: How to bridge the gap between clinical and pre-clinical perspectives. *Brain, Behavior, and Immunity,* ***26***, 1169–1179.

[11] Hovens, I.B., Schoemaker, R.G., van der Zee, E.A., Absalom, A.R., Heineman, E., van Leeuwen, B.L. (2014). Postoperative cognitive dysfunction: involvement of neuroinflammation and neuronal functioning. *Brain, Behavior, and Immunity,* ***38***, 202-210.

[12] Hovens, I.B., Schoemaker, R.G., van der Zee, E.A., Heineman, E., Nyakas, C., van Leeuwen, B.L. (2013). Surgery-induced behavioral changes in aged rats. *Experimental Gerontology,* ***48***, 1204–1211.

Video analysis of social interactions in mice: An integrated solution

F. de Chaumont[1], S. Granon[2] and J.-C. Olivo-Marin[1].

[1]Bio Image Analysis Unit, CNRS URA 2582, Institut Pasteur, France
chaumont@pasteur.fr
[2] Université Paris Sud 11, CNRS UMR 8195, Orsay, France.

Prior to any analysis, the entry point for a behavioral study is the tracking of one or several animals. One of the main issues when dealing with more than one animal is the fact that they do have frequent contact and are difficult to individualize.

To recognize the identity of each animal, several methods can be used such as shaving or painting the animals. In both cases, this may bias the experiment because of either the social meaning of being shaved (stress or dominance) or the smell of the painting. In our solution [1], we chose as a constraint not to change the appearance of animals.

In the described method, we introduce physical models, built by assembling basic physical primitive links together by articulations (i.e. degrees of freedom) created by either elastic wires or physical joints (rotation, sliders). To perform the tracking [2], the physical model is adjusted to the underlying image data thanks to forces that are computed from the image data and are then applied to the virtual model in order to adjust it to the location of the animal. The added robustness given by the model allows the tracking of several mice, and provides the location of the head, the belly, and the beginning of the tail for each animal.

We created a repertoire of events based on behaviorist's knowledge which describes each event in a mathematical manner, thus making the labeling reproducible.

Different kinds of event can be measured: Single frame events such as 'head-head contact', or events occurring on a number of successive frames as it is the case for pursuits. The repertoire also introduces the labelling of complex temporal sequences of events which represent the combination of several simpler events such as 'Mouse A comes to mouse B, contact, and then the contact is stopped by Mouse B.' Thanks to the tracking and the repertoire, video sequences are automatically analyzed and events are labeled [3].

The software implementation allows one to watch and browse the entire chronogram of events. The user can seek in the video and see information in overlay at the same time:

1. The quality of the tracking, as the head, belly and beginning of the tail are displayed.
2. The field of vision of the mice.
3. The different events that are occurring at the current frame.

The final part of the software [4] allows one to compile chronograms to create statistical transition graphs, which describe what are the predecessor and successor of each event. Those graphs are compiled for each strain and sum up all the data in a single view.

All the software [2-4] is free and open source, it can be downloaded through the software Icy [5]. A complete video showing how to use the software is also available [6].

Ethical statement

The mice were treated according to the ethical standards defined by the Centre National de la Recherche Scientifique for animal health and care in strict compliance with the EEC recommendations (86/609).

References

[1] de Chaumont, F., Coura, R.D.S., Serreau, P., Cressant, A., Chabout, J., Granon, S., Olivo-Marin, J.-C. (2012). Computerized video analysis of social interactions in mice. *Nature Methods.* ***9****(4)*, 410–417.

[2] de Chaumont, F., et al. (2012). http://icy.bioimageanalysis.org/plugin/Mice_Profiler_Tracker

[3] de Chaumont, F., et al. (2012). http://icy.bioimageanalysis.org/plugin/Mice_Profiler_Video_Label_Maker

[4] de Chaumont, F., et al. (2012). http://icy.bioimageanalysis.org/plugin/Mice_Profiler_Label_Analyser

[5] Icy software. http://icy.bioimageanalysis.org

[6] https://www.youtube.com/watch?v=9sgoB-StV-k

JAABA: An interactive machine learning system for automatic annotation of animal behavior

K. Branson[1], A. A. Robie[1], M. Kabra[1], M. Rivera-Alba[1], S. Branson[1,2]

[1]HHMI Janelia Farm Campus, Ashburn, VA, USA

bransonk@janelia.hhmi.org

[2]Electrical and Computer Engineering, Caltech, Pasadena, CA, USA

We present the Janelia Automatic Animal Behavior Annotator (JAABA), a new machine learning-based system to enable researchers to automatically compute highly descriptive, interpretable, quantitative statistics of the behavior of animals [1]. Through our system, the user encodes their intuition about the structure of behavior in their experiment by labeling the behavior of the animal, e.g. walking, grooming, or following, in a small set of video frames. Our system uses machine learning techniques to convert these manual labels into behavior detectors that can then be used to automatically classify the behaviors of the animals in a large data set with high throughput. We combine an intuitive graphical user interface, a fast and powerful machine learning algorithm, and visualizations of the classifier into an interactive, usable system for creating automatic behavior detectors.

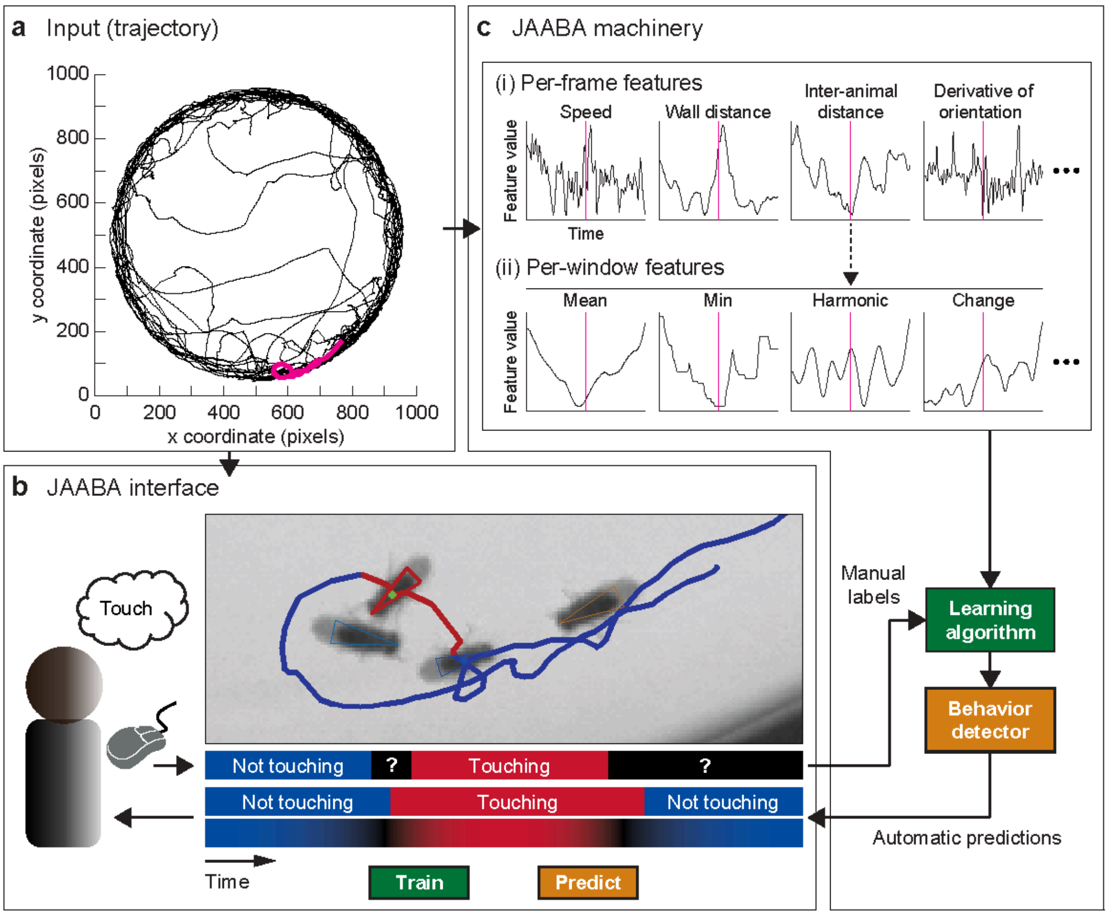


Figure 1. JAABA overview. (a) Input trajectory (x,y position over 1000s). (b) JAABA interface. The top time line shows the user's manual labels, and the bottom two time lines show the classifier's predictions and confidence. (c) JAABA machinery. (i) Example "per-frame" feature time series. (ii) Example "window" feature time series.

JAABA operates on the animals' trajectories computed from automatic tracking methods (Figure 1a). From these, we compute simple “per-frame” features such as instantaneous speed or distance to the closest animal (Figure 1c(i)). Next, we compute “window features” that describe the distribution of per-frame features around the current frame (Figure 1c(ii)). We compute the mean, standard deviation, minimum, etc. for multiple window sizes and temporal offsets. All window features can be computed efficiently using convolution or image morphology. We compute between 5,000 and 10,000 window features to represent each frame for each animal. We use an instantiation of Boosting (GentleBoost), modified for speed, to train classifiers. Training takes 15-45 seconds, depending on the data set.

We demonstrate that our system can be used by scientists without expertise in machine learning to independently train accurate behavior detectors. We show that it is sufficiently powerful and general-purpose to train a large, diverse set of single-animal and social behavior classifiers for flies (15 behaviors), mice (2 behaviors), and *Drosophila* larvae (3 behaviors), achieving an average per-frame error rate of 2.7% and a maximum error rate of 6% over all behaviors, compared to human annotations. We show that JAABA is usable by non-computer scientists: we gave 12 volunteers a 15-minute presentation on how to install and use JAABA, then asked them to each train a “chase” classifier for flies. The error rates achieved were 2% on average and 5% maximum over users. Finally, we show that JAABA can be used to create behavior classifiers robust enough to successfully be applied to a large, phenotypically diverse data set consisting of thousands of transgenic lines of *Drosophila melanogaster* collected in our thermogenetic screen of the Janelia GAL4 collection. Statistics of the automatic behavior classifications such as the fraction of time spent performing a given behavior are powerful descriptions, and we show that these statistics can be used to map neuronal anatomy to behavior in this neural activation screen.

Our system is complementary to video-based tracking methods, and we envision that it will facilitate extraction of detailed, scientifically meaningful measurements of the behavioral effects in large experiments.

Reference

[1] Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., Branson, K. (2013). JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods, **10***, 64-67.

Automated discovery of behavioural patterns in rodents

Diego Sona[1], Matteo Zanotto[1], Francesco Papaleo[2], Vittorio Murino[1]

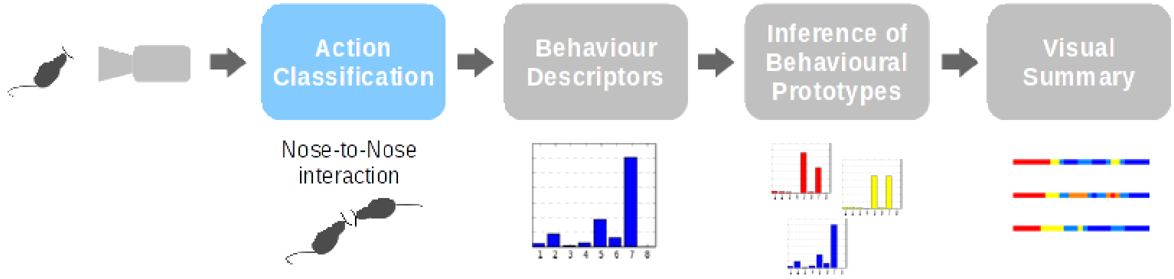
[1]Pattern Analysis and Computer Vision – Istituto Italiano di Tecnologia – Genova – Italy
[2]Neuroscience & Brain Technologies – Istituto Italiano di Tecnologia – Genova – Italy
<name>.<lastname>@iit.it

Genetically modified mouse models are now routinely assessed for their behavioural phenotype, as they constitute a valuable tool for translational studies investigating social abnormalities in mental illnesses. Reliable analysis of complex social behaviours comes through the adoption of objective and statistically significant indicators, requiring therefore measures taken over extended time spans. In such investigations, manual scoring of social interactions is still the most important experimental bottleneck, since it is extremely challenging and time consuming. Indeed, long-lasting investigations eliciting long-term behavioural patterns are impractical, and the experimenter is pushed to assess the theories with manageable experiments necessarily short and of limited extent. Moreover, manual scoring is always prone to the risk of subjective judgments due to the experimenter bias, with high impact on repeatability of experiments.

Hence, there is an increasing interest in the development of systems for automated behaviour analysis from videos, enabling long lasting and large-scale experiments. While this opens huge opportunities for high throughput phenotyping, the increasingly wide quantity of collected data force the scientists to face new challenges for their analysis, especially when relevant information cannot be captured by mere statistics. This is often the case when dealing with complex phenomena such as relations, sequences, co-occurring events or repetitive patterns, especially if they are evolving over time.

Aiming at the above issues, we developed a computational framework combining a tracking algorithm able to simultaneously track multiple mice and an automatic system for the analysis of social interactions of multiple interacting mice using a learning-by-examples approach. Tracking was addressed by developing a robust processing pipeline, composed of blob detection, shape segmentation and matching modules, allowing to address the frequent occlusions of mice, which have non-rigid shapes and similar appearance. Tracking results are then used by an automatic behavioural classification algorithm, based on an extension of random forests to the multi-frame case. This multi-frame extension allows exploiting the temporal information at the level of the model rather than at the feature level, in order to recognize behaviours that can be inferred only considering a certain time span. The system is characterized by a great adaptability to the experimenter needs, who can refine existing classes of interaction or create new ones just providing examples, i.e., annotating the behaviours in segments of video. In our system the social interactions are detected on all possible mice pairs and then properly combined to obtain overall behaviours for each mouse. This permits either global analyses in the arena or finer analysis over a specific mouse in a group.

Mice behaviour phenotyping, however, is not only made of mere statistics. While the common trend is to focus on statistics of single actions, sequences of actions and repetitions of behaviours are important elements to characterize behavioural phenotypes. Indeed, looking for patterns of actions can give a richer interpretation of the observed behaviours. For this reason, we propose a mathematical formalism based on a Bayesian Nonparametric model (a Dirichlet Process Mixture of Multinomials) able to capture an appropriate collection of behavioural prototypes in a completely unsupervised way. These prototypes are then used to characterize mice behaviour and to present the user with a high-level aggregation and an intelligible representation of the statistical structure of behaviour, capturing relevant behavioural patterns not known a priori.



The proposed method aims at supporting behavioural scientists in the complex task of understanding important traits of mice behaviour. Indeed, such behaviour-oriented data mining approach could be very useful to get a first understanding of the data, providing new chances for the identification and interpretation of novel complex behavioural traits, and can guide the formulation of hypothesis and the subsequent design of further experiments.

Temporal structure of rat behavior in the social interaction test

M. Casarrubea[1,4]*, D. Cancemi[3], A. Cudia[3], F. Faulisi[1], F. Sorbera[1], M.S. Magnusson[2], M. Cardaci[3,4], G. Crescimanno[1,4]

(1) - Dept. of Experimental Biomedicine and Clinical Neurosciences (BioNec), Human Physiology Section "G. Pagano", Laboratory of Behavioral Physiology, University of Palermo - Palermo, Italy
(2) - Human Behavior Laboratory, University of Iceland - Reykjavik, Iceland
(3) - Dept of Psychological Sciences, Pedagogy and Formation, University of Palermo - Palermo, Italy
(4) - Interdepartmental Center of Knowledge Technologies (CITC)
* maurizio.casarrubea@unipa.it

Introduction

The social interaction test is an important tool to study anxiety in rodents. Basically, the rationale of this test is that an increase of the interaction is indicative of a reduced anxiety; conversely, a decreased interaction indicates an increased anxiety [6]. Thus, it is not surprising that a common application of the social interaction test is the evaluation of the behavioral effects induced by anxiety modulators. However, despite the high number of studies in which the social interaction test is used, surprisingly scanty data are available on the temporal structure of the behavior of two interacting rodents. On this subject, several interesting questions and topics of discussion still remain unanswered. For instance: whether the behavior of the pair has a temporal organization, that is, if specific behavioral sequences occur during the interaction of the two subjects; second, if a temporal organization exists, could it be possible to discuss such an organization in ethological terms? Could it be possible to identify behavioral sequences with a specific functional meaning? This study has been designed to shed light on the aforementioned questions. To this aim, a refined analysis of the temporal features of the behavior, known as t-pattern analysis, has been applied. This analysis is a multivariate approach developed to determine whether two or more behavioral events occur sequentially and with significant constraints on the interval lengths separating them [10].

Method

Eighteen male Wistar rats, two months old were used. Animals were housed in a thermo-regulated room maintained at constant temperature (23±1 °C), under a 12 h light/dark cycle (lights on = 07:00 a.m.; lights off = 07:00 p.m.). Food pellets and water were provided ad libitum. Each subject has been individually housed for 5 days before testing [5] [6] [7]. On the test day, to minimize transfer effects and avoid possible visual or olfactory influences, two rats, randomly taken from housing room, were transferred inside their own home cages to testing room and allowed to acclimate for 30 min far from the experimental apparatus. Environmental temperature in testing room was maintained equal to the temperature measured in the housing room. All rats, experimentally naïve, were observed only once. Each pair was observed for 15 minutes in a 50 x 50 cm open field apparatus. The behavior was recorded by means of a digital video-camera and video-files stored in a backup device for following analyses. Table 1 presents the utilized ethogram, arranged on the basis of previous studies of different Authors [11] [13] [14] [15]. All video files were coded using The Observer (Noldus Information Technology, The Netherlands).

Table 1. Ethogram of rat behavior in the social interaction test.

Intra-subject	Abbr.	Description
Walking	Wa	The rat walks around sniffing the environment
Climbing	Cl	The rat maintains an erect posture leaning against the Plexiglas wall. Usually associated with sniffing.
Rearing	Re	The rat maintains an erect posture without leaning against the wall. Usually associated with sniffing.
Front Paw Licking	FPL	The rat licks or grooms its forepaws
Hind Paw Licking	HPL	The rat licks or grooms its hind paws
Face Grooming	FGr	The rat rubs its face with the forepaws
Body Grooming	BGr	The rat rubs the body combing the fur by fast movement of the incisors
Shaking	Sh	The rat shakes its head and body with rapid semicircular movements
Immobility/Resting	Imm	The rat maintains a fixed posture
Immobile Sniffing	ISn	The rat sniffs the environment standing on the ground
Inter-subject	**Abbr.**	**Description**
Withdrawing	Wit	One rat walks or runs away from the other rat.
Following	Fol	One rat follows the partner while the other is walking away
Approaching	App	One rat walks in the direction of the partner, while the other rat is immobile or is alredy approaching him
Crawling over	CrO	One rat walks over the partner
Crawling under	CrU	One rat walks under the partner
Boxing/Wrestling	Box	Offensive/aggressive behaviors such as pawing, pouncing, nosing, biting, boxing, kicking, wrestling
Leaning on	LeO	One of the rats leans with its forelimbs on the other rat that, in turn, maintains all the four paws on the ground.
On-top	Top	One of the animals stands over the partner that lies with its back on the floor
On-back	Bck	One of the animals lies with its back on the floor with the other animal standing over it
Mounting	Mnt	One of the rats holds the other rat's trunk with the forelimbs
Social grooming	SoG	One of the rats grooms the partner's body, neck or face
Social sniffing	SoS	One of the rats sniffs the partner's face and/or body
Ano-genital sniffing	GeS	One of the rats sniffs the partner's anogenital area

To assess the temporal relationships among behavioral events, event log files, obtained from coding process, were analyzed by means of the software Theme (Noldus Information Technology bv, The Netherlands; Patternvision Ltd, Iceland). The software can analyze any time-based data; on this subject it is important to mention that the time unit can be of any size (e.g. seconds, milliseconds, days, years), depending on the system being studied. Theme's detection algorithm searches for statistically significant relationships between events in behavioral data, by taking into account the order, timing, and frequency of these events [10]. In brief, given an observational period T0-Tx where several behavioral events occur, the algorithm compares the distributions of each pair of the behavioral elements "A" and "B" searching for an interval so that, more often than chance expectation, A is followed by B within that time window. In this case A and B do represent a two-element t-pattern, indicated with the string (A B). In a second step, such first level t-patterns are considered as potential terms in higher patterns, for example, ((A B) C) and so on. Thus, following this hierarchical bottom-up detection process, more complex patterns may be identified up to any level. Before a t-pattern search is performed, the software requires specific search parameters which, in the present study, are: "significance level" (maximum accepted probability of any critical interval relationship to occur by chance) = 0.0001; "lumping factor" (forward and backward transition probability above which A and B of a t-pattern (A B) are lumped, that is, A and B are not considered separately but only as the (A B) pattern = 0.90; "minimum samples" percent of subjects in which a pattern must occur to be detected) = 100; "minimum occurrences" (minimum number of times a t-pattern must occur to be detected) = 9. Such parameters were selected to obtain the identification of behavioral sequences that were present in all the samples and were characterized by high level of significance. More exhaustive descriptions of concepts, theories and methodological approaches underlying t-pattern analysis can be found in previous works [1] [2] [3] [10].

Results

221 different t-patterns have been detected. On the basis of the quality of the events present in the structure of these patterns (namely, intra- or inter-subject events, see tab. 1), four different categories have been identified (fig. 1): a first category consists of t-patterns containing exclusively inter-subject events; a second one consists of t-patterns containing both inter- and intra-subject events; a third category encompasses patterns with events performed both by subject 1 and subject 2 as well; finally, a category consists of patterns with events performed by one of the two subjects. Figures 2-5 illustrate the terminal strings and the occurrences/onset of the most complex and frequent t-pattern for each category.

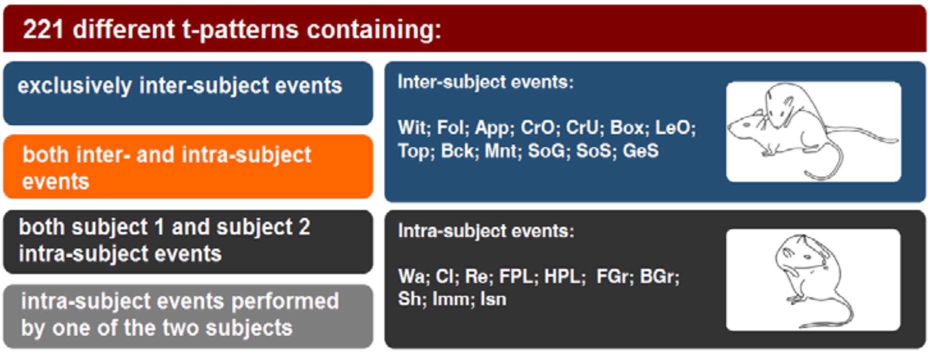


Figure 1. T-pattern analysis has revealed 221 different t-patterns. Left: four different categories of t-patterns; right: structural events (inter- and intra-subject) from the ethogram. For abbreviations see tab. 1.

Discussion

Multivariate analysis has revealed four different sets of t-patterns (figure 1): t-patterns where the sequence consists exclusively of structural components defined on the basis of the relationships with the partner, namely, inter-subject components (figure 1 and figure 2). A t-pattern encompassing only this type of events is indicative of the interaction between the two tested subjects.

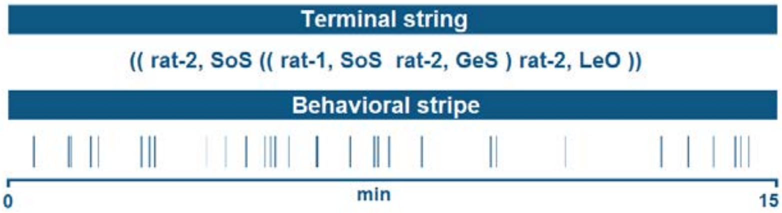


Figure 2. Terminal string and behavioral stripe of the most complex and frequent t-pattern for the category containing exclusively inter-subject events. For abbreviations see tab. 1.

A second category of t-patterns encompasses sequences consisting both of inter- and intra-subject events (figure 1 and figure 3). For the definition of what a t-pattern is (namely, a statistically significant sequence of events), all the patterns containing inter-subject components do represent an interaction as well. Our analysis has also detected various t-patterns consisting only of intra-subject events. Such an outcome calls for an interesting question/topic of discussion: is it correct to consider a t-pattern containing intra-subject events, carried out by the two rats, representative of an interaction? For instance, does the sequence ((rat-2,Wa rat-2,Cl)(rat-1,ISn rat-1,Wa)) illustrated in figure 4 represent an interaction?

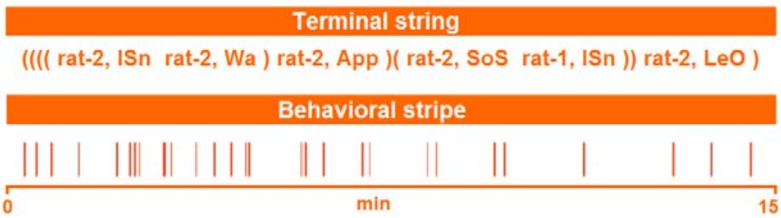


Figure 3. Terminal string and behavioral stripe of the most complex and frequent t-pattern for the category consisting both of inter- and intra-subjects events. For abbreviations see tab. 1.

The answer should be affirmative because such a t-pattern reveals a clear-cut causality between the behaviors of the two subjects, namely, following Walking (Wa) and Climbing (Cl) produced by the rat-2, rat 1 produces

Immobile-Sniffing (ISn) and Walking (Wa); importantly, such a succession of events is significantly repeated several times along the 0 - 15 min observational period, as clearly illustrated by the corresponding stripe (figure 4).

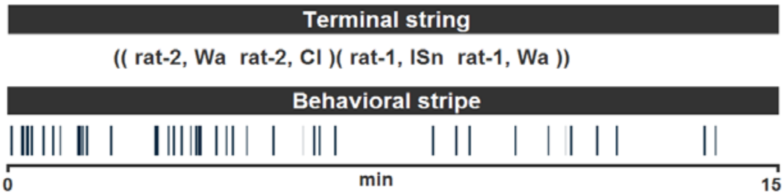


Figure 4. Terminal string and behavioral stripe of the most complex and frequent t-pattern for the category containing intra-subject events carried out by the two rodents. For abbreviations see tab. 1.

Finally, it has been possible to identify a fourth set of t-patterns containing intra-subject events produced by one of the two animals. In this context, the interaction is absent because the rat carries out a behavioral sequence consisting of events totally disjointed from the partner both from a qualitative and temporal point of view. Figure 5 exemplifies such a fourth category: the whole sequence is carried out by rat-2 and does contain exclusively Immobile-Sniffing (ISn), Walking (Wa) and Rearing (Re), that is, intra-subject activities.

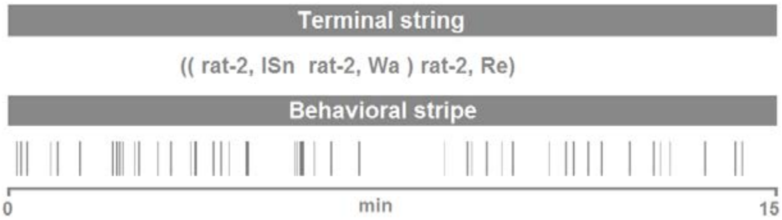


Figure 5. Terminal string and behavioral stripe of the most complex and frequent t-pattern for the category containing only intra-subject events carried out by one of the two subjects. For abbreviations see tab. 1.

These preliminary results demonstrate that rat behavior in the social interaction test is structured on the basis of several recurring sequences of events characterized by statistically significant constraints on the interval lengths separating them. The multivariate t-pattern analysis could represent a valid approach to study hidden characteristics of rodents social behavior; an example, on this subject, is represented by the detection of t-patterns containing only intra-subject components where a clear, otherwise undetectable, causality in the behavioral sequences of the two animals can be identified. Finally, t-pattern analysis, coupled with refined approaches aimed at the automated analysis of behavior [4] [8] [9] [12], could represent a valuable tool to investigate the temporal features of social behavioral dynamics, anxiety-related behaviors and, as a consequence, to assess the efficacy and/or the effects induced by the administration of potential anxiety modulators.

Ethical Statement

All efforts were made to minimize the number of animals used and their suffering. All the experimental procedures were conducted in accordance with the European Communities Council Directive (86/609/EEC) and approved by the Veterinary Committee officially appointed by the University of Palermo.

References

[1] Casarrubea, M., Sorbera, F., Magnusson, M.S., Crescimanno, G. (2011). T-pattern analysis of diazepam-induced modifications on the temporal organization of rat behavioral response to anxiety in hole-board. *Psychopharmacology*, ***215***, 177-189.

[2] Casarrubea, M., Roy, V., Sorbera, F., Magnusson, M.S., Santangelo, A., Arabo, A., Crescimanno, G. (2013). Temporal structure of the rat's behavior in elevated plus maze test. *Behavioural Brain Research*, ***237***, 290-299.

[3] Casarrubea, M., Roy, V., Sorbera, F., Magnusson, M.S., Santangelo, A., Arabo, A., Crescimanno, G. (2013). Significant divergences between the temporal structure of the behavior in Wistar and in the spontaneously more anxious DA/Han strain of rats tested in elevated plus maze. *Behavioural Brain Research*, ***250***, 166-173.

[4] de Chaumont, F., Coura, R.D., Serreau, P., Cressant, A., Chabout, J., Granon, S., et al. (2012). Computerized video analysis of social interactions in mice. *Nature Methods*, ***9(4)***, 410-417 - doi:10.1038/nmeth.1924.

[5] Dekeyne, A., Denorme, B., Monneyron, S., Millan, M.J. (2000). Citalopram reduces social interaction in rats by activation of erotonin (5-HT)2C receptors. *Neuropharmacology*, ***39***, 1114-1117.

[6] File, E.S., Seth, P. (2003). A review of 25 years of social interaction. *European Journal of Pharmacology*, ***463***, 35-53.

[7] File, S.E., Lippa, A.S., Beer, B., Lippa, M.T. (2004). Animal tests of anxiety. *Current Protocols Neuroscience* - DOI: 10.1002/0471142301.ns0803s26.

[8] Giancardo, L., Sona, D., Huang, H., Sannino, S., Managò, F., Scheggia, D. et al. (2013). Automatic visual tracking and social behaviour analysis with multiple mice. *PLoS ONE*, ***8(9)*** - DOI:10.1371/journal.pone.0074557.

[9] Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., Branson, K. (2013). JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, ***10(1)***, 64-67.

[10] Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: t-patterns and their detection. *Behavior Research Methods, Instruments, & Computers*, ***32***, 93-110.

[11] Spruijt, B.M. (1992). Progressive Decline in Social Attention in Aging Rats: An Information-Statistical Method. *Neurobiology of Aging*, ***13***, 145-151.

[12] Spruijt, B.M, Peters, S.M., de Heer, R.C., Pothuizen, H.H., van der Harst, J.E. (2014). Reproducibility and relevance of future behavioral sciences should benefit from a cross fertilization of past recommendations and today's technology: "Back to the future". *Journal of Neuroscience Methods* - DOI: 10.1016 / j.jneumeth.2014.03.001.

[13] Van Den Berg, C.L., Van Ree, J.M., Spruijt, B.M. (1999). Sequential analysis of juvenile isolation-induced decreased social behavior in the adult rat. *Physiology & Behavior*, ***67***, 483-488.

[14] Vanderschuren, L.J., Spruijt, B.M., Hol, T., Niesink, R.J., Van Ree, J.M. (1995). Sequential analysis of social play behavior in juvenile rats: effects of morphine. *Behavioural Brain Research*, ***72***, 89-95.

[15] Yamada-Haga, Y. (2002). Characteristics of social interaction between unfamiliar male rats (Rattus norvegicus): comparison of juvenile and adult stages. *Journal of Ethology*, ***20***, 55-62.

Automated Classification of Rat Social Behavior

S.M. Peters [1,2], I.J. Pinter[2], R.C.de Heer[2], J.E. van der Harst[2], B.M. Spruijt[1]

[1]Department of Behavioral Biology, Faculty of Science, Utrecht University, the Netherlands
s.m.peters@uu.nl
[2]Delta Phenomics B.V., Schaijk, the Netherlands

Introduction

Several psychopathologies affecting social behavior such as autism disorders, major depression disorder and schizophrenia are detrimental diseases for which currently no cure exists and much of the pathology is still unknown. The development of valid animal models and reliable test paradigms are crucial elements in the process of understanding the underlying mechanisms of these disorders. Rodents are, similar to humans, very social animals. This aspect, however, has been neglected in behavioral neuroscience and drug discovery. While the quantification of rodent social behavior has long relied on manual observations, recent advanced software and/or hardware technologies e.g. [1-5] make it possible to automatically monitor multiple animals and their social interactions. This requires a different approach in analyzing the data that is generated by these systems. Here, we propose a method to analyze social interactions of rat pairs. Since this method is not relying on subjective human interpretation, but on computerized data, it is more objective and less labor intensive. Also, we will argue here that our method is very valid in ethological terms, because criteria differentiating various behaviors are based on the occurrence (using individual frequency distributions) of the animal's own behavior.

Methods

Pairs of male rats of 5 weeks old were placed in a relatively large PhenoTyper® cage of 90 cm x 90 cm (Noldus Information Technology, the Netherlands) and allowed to freely interact (i.e. social interaction test). At this age rats are known to display the highest levels of social (play) behavior (e.g. [6, 7]). All testing was done during the dark phase (that is the active phase) of the animals under red light conditions and all animals were habituated to human handling. Before each social interaction test, animals were marked red or black (permanent marker, Edding, Germany) for the software to recognize both individuals based on their marking. The black marking is visible in the video, while the red marking is not visible (due to the infrared light conditions, see below). In this way both animals experience the same practical procedure, but the software recognizes a marked and an unmarked animal; a similar procedure is used by [8, 9]. Incorrect identity swops (in the initial version of the software this was about 50% of the samples) made by the software were manually corrected after video tracking. The experiments were performed in adherence to the legal requirements of Dutch legislation on laboratory animals (Wod/Dutch 'Experiments on Animals Act') and were reviewed and approved by an Animal Ethics Committee ('Lely-DEC').

From each social interaction test session, top view video recordings were made with a top-unit placed on top of the PhenoTyper, containing an infrared sensitive camera (CCD 1/3" SONY SUPER HAD CCD black/white) and IR-filter (type Kodak 87C). In the development of our method we were inspired by Golani and coworkers that have for long been working on automatic classification of rodent behavior, e.g. [10]. Their approach involves the Gaussian expectation maximization method to search for natural modes in the data based on frequency distributions (see [11] for an extensive explanation of this method). After tracking the animal from video recordings with automatic tracking software (EthoVision XT 8.0, Noldus Information Technology, the Netherlands), raw data containing x- and y-coordinates was exported and further analyzed in MatLab® R2012b (The MathWorks, United States). First, raw track data was smoothed using a robust LOWESS filter with a 1-s time window in order to remove noise. Then, parameters such as the distance moved, speed of movement and distance between two animals were determined. The tracking data was divided between movement bouts and stops (i.e. arrests). Therefore, data was filtered with a repeated running median using a one dimensional median filter with order 7, 5, 3 and 3 respectively. Based on visual inspection of the tracks of the animals and video images the threshold for a stop was set at 0.07 cm between 2 samples lasting at least 0.16 seconds. Next, a density estimator (i.e. smoothed frequency distribution) of the maximal velocity of each movement bout was made. On these density estimator plots the best Gaussian curves to represent the empirical data were fitted using an expectation maximization (EM) method. Subsequently, the intersection of these curves determined the different “modes” or velocity categories in which the animals moved. Additionally, the same analysis was done on the distance between the animals, with the difference that this parameter was calculated for each sample in the track data. Also, for this distance between parameter the intersections of the Gaussian curves that best represented the empirical data were used to determine the different “modes” of proximity. Thereafter, we combined both movement and proximity modes together to obtain specific approach and avoidance behaviors characterized by velocity and distance.

Results and Discussion

Our method revealed that in our setup the movement of juvenile rats (when freely socially interacting) can be divided between movements with either low velocities or high velocities (average threshold of 34 cm/s). In addition, we found that distance between rat pairs can be classified in 3 different proximities categories: in -very-close proximity, in proximity and not in proximity (average thresholds: 11 and 31 cm, respectively). Combining these two modes (velocity with proximity) we could clearly identify for example behavioral bouts in which the animals are moving with high velocity and in close proximity, representing chasing behavior. Currently, we are also investigating social interactions of adult rat pairs with this method and we are applying our method on existing models used to examine social deficits, for example phencyclidine induced social impairments.

Using automated social parameters based on coordinates of the animals is not completely new. Sams-Dodd (1995) used a similar method to create the automated parameter ‘social interaction’ [8]. This parameter was made with a fixed threshold; animals were regarded to be in proximity when their center of gravity points are within 20 centimeters from each other. He mentioned that: “selection of this criterion value of 20 cm is based on systematic variation of this parameter from 0 to 50 cm” and “the value of 20 cm resulted in the least variation in the data”. It is exactly this variation that we now use to classify the different categories of proximity. Important benefit of this approach is that the threshold is not created artificially by limiting variation, but based on the animal’s own behavior and natural variation within this behavior. In addition, it is now also possible to combine the information represented by speed of movements of the animals with the categories of proximity.

The next step is to apply this method on data from continuously monitored group-housed rats to study their social behavior displayed in a home-cage environment for more precise registration and thus enhanced differentiation of rodent social behavior. This is also beneficial for increasing the translational value of social behavior when employed as an indicator for assessing treatment efficacy in animal models for psychopathology involving social behavior.

Acknowledgements

The large PhenoTyper cage was developed as part of the SenseWell project. We acknowledge our partners in this consortium (Metris BV, Noldus Information Technology BV, Telemetronics Biomedical BV, BioMetris Wageningen University, Faculty of Veterinary Medicine Utrecht University and EEMCS TU Delft) for their support. This research was partially financed by a grant from Agentschap NL (NeuroBasic-PharmaPhenomics) and ICT Regie / NWO (SenseWell).

References

[1] de Chaumont, F., Coura, R.D., Serreau, P., Cressant, A., Chabout, J., Granon, S., Olivo-Marin, J. (2012). Computerized video analysis of social interactions in mice. *Nature Methods*, ***9****, 410-417.*

[2] Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., Branson, K. (2013). JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, ***10***, 64-67.

[3] Giancardo, L., Sona, D., Huang, H., Sannino, S., Managò, F., Scheggia, D., Papaleo, F., Murino, V. (2013). Automatic Visual Tracking and Social Behaviour Analysis with Multiple Mice. *PLoS ONE*, ***8****(9)*, e74557.

[4] Weissbrod, A., Shapiro, A., Vasserman, G., Edry, L., Dayan, M., Yitzhaky, A., Hertzberg, L., Feinerman, O., Kimchi, T. (2013). Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nature Communications*, ***4***, 2018.

[5] Shemesh, Y., Sztainberg, Y., Forkosh, O., Shlapobersky, T., Chen, A., Schneidman, E. (2013). High-order social interactions in groups of mice. *eLife*, ***2***, e00759.

[6] Panksepp, J. (1981) The ontogeny of play in rats. *Developmental Psychobiology*, ***14***, 327-332.

[7] Pellis, S.M., Pellis, V.C. (2007). Rough-and-tumble play and the development of the social brain. *Current Directions in Psychological Science*, ***16***, 95-98.

[8] Sams-Dodd, F. (1995). Automation of the social interaction test by a video-tracking system: Behavioural effects of repeated phencyclidine treatment. *Journal of Neuroscience Methods*, ***59***, 157-167.

[9] Spruijt, B.M., Hol, T., Rousseau, J. (1992). Approach, avoidance, and contact behavior of individually recognized animals automatically quantified with an imaging technique. *Physiology & Behavior*, ***51***, 747-752.

[10] Benjamini, Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., Golani, I. (2010). Ten ways to improve the quality of descriptions of whole-animal movement. *Behavioral Brain Research*, ***34***, 1351-1365.

[11] Drai, D., Benjamini, Y., Golani, I. (2000). Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods*, ***96***, 119-131.

Laban movement analysis for action recognition

A. Truong, H. Boujut, T. Zaharia

ARTEMIS Department, Institut Mines-Telecom, Telecom SudParis, CNRS UMR 8145 – MAP5, France.
{arthur.truong, hugo.boujut, titus.zaharia}@telecom-sudparis.eu

Abstract

In this paper, we introduce a new 3D expressive model of gesture descriptors based on the Laban Movement Analysis (LMA). The proposed model is tested and evaluated within an action recognition context. Experimental results, obtained on the Microsoft Research Cambridge-12 dataset, show that the approach yields very high recognition rates (more than 97%).

Introduction

The interpretation of gestures is a stake for numerous applications: e-health, video games, artistic creation, video surveillance... It involves computer vision, machine learning methodologies, and requires the elaboration and development of effective gesture descriptors. Throughout the last decade, several researchers investigated the field of gesture/posture/action recognition. Let us quote some approaches based upon global representation like body angles shaped by limbs positions [1] or spatio-temporal "bouding volumes" [2]. Other approaches are based upon local representations of motion by points saliency, which can be evaluated according to the information contained in spatiotemporal neighborhood [3] or computed in relationship with energy functions putting at stake color, intensity and orientation conspicuities [4].

These different works fail to propose effective descriptors models, and barely take into account the expressive character of body gestures in their analysis. In this context, a key issue concerns the modeling of the various features involved in gestures inter-subjective character, and the testing of these features ability to characterize different types of actions. After its validation, such a general gesture model could then be used for the analysis of other types of high-level contents, like emotions or affects. In this paper, we propose a set of body motion expressive descriptors, inspired from the LMA model [5] and test their relevance in action recognition context.

Proposed approach

The Laban Movement Analysis (LMA [5]) model, introduced by Rudolf Laban consists of describing body movement in terms of qualities relating to different characterizations of the way this movement is performed. More precisely, the LMA model is composed of the following five major qualities: *Body*, *Space*, *Effort*, *Shape*, and *Relationship*. For this study, we only considered the *Space*, *Effor*t and *Shape* components, for they relate to expressivity and communicational aspects (The *Body* quality deals with the structural aspect of the gesture and the *Relationship* quality is better suited for group analysis). The *Space* component refers to the place, direction and path of the movement. The *Effort* component depicts how the body concentrates its effort to perform the movement and deals with expressivity and style. The *Effort* further includes the following sub-components: *Space* (not to be confused with *Space* quality) that we did not retain in our work because of its redundancy with the generic *Space* component; *Time* which separates movements between sudden and sustained (or continuous) ones; *Flow* which describes movements as free or constrained; *Weight*, which aims at distinguishing between light and heavy movements. The *Shape* component includes 3 factors: *Shape Flow*, which relates to participant's concern about changing relationships between its different body parts; *Directional movement*, which describes the direction of the movement toward a particular point; *Shaping*, characterizing how the body changes its shape in a particular direction: rising/sinking, retreating/advancing and enclosing/spreading, which will be retained for our model for it is the most concrete *Shape* sub-quality.

Some first studies like [6] attempted to identify and classify cetain Laban qualities, based on visual gesture descriptors, with the help of supervised classification approaches, showing that a mid-level Laban representation can be obtained starting from visual descriptors. Our objective is slightly different: define a set of descriptors characterizing the individual Laban qualities that we considered relevant (i.e. *Space*, *Effort* and *Shape*), and exploit them directly for action recognition purposes, without explicitly determining the Laban qualities.

The proposed descriptors are associated to 3D body skeleton joints trajectories that can be recorded with a depth sensor (*e.g.*., Kinect camera) at a rate of 30 frames per second. The Kinect sensor provides a maximum number of 20 joints (Figure 1.a). For each body joint *i*, the trajectory will be represented as a sequence $(x_{i,t}, y_{i,t}, z_{i,t})_{t=0}^{N-1}$ where N is the total number of frames. For all the joints, $(x_{i,t}^{trans}, y_{i,t}^{trans}, z_{i,t}^{trans})_{t=0}^{N-1}$ will refer to a trajectory transform consisting in moving the body at every time t so that: a) the shoulders and the hip center belong to a same plane parallel to (yOz) plane; b) both shoulders are at the same height. Figure 1.b illustrates the result of this body alignment process, so that (xOy), (yOz) and (zOx) planes respectively correspond to sagittal, vertical and horizontal body planes. The approach leads to a total number of 81 features we are now introducing.

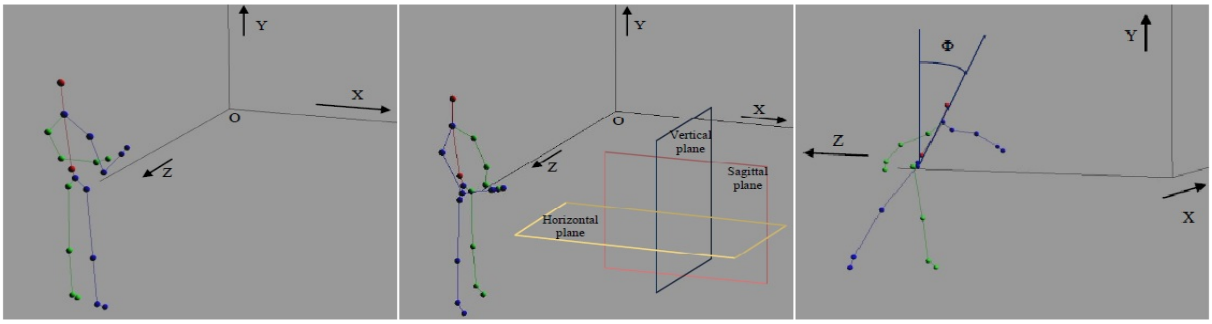


Figure 1. Body skeleton joints at a particular frame (a); skeleton joints new positions after transforms application and body planes representation (b); illustration of forward tilt angle (c).

For the *Space* quality (9 features), the first characteristic is the total length of the head trajectory. Then, we retain the number of zero crossings *($n_{ZC}(x'^{trans}_{Head,t})_{t=0}^{N-1}$))* of the first derivative of the head's component in the direction parallel to the vertical plane (which measures the number of head's retreats/advances), the maximal amplitude of the head movement following this direction $(max_{t=0}^{N-1}(|x_{Head,t}^{trans}|) - min_{t=0}^{N-1}(|x_{Head,t}^{trans}|))$ and the relative temporal instant of maximum's reaching: t_{max}/N. Then, we consider the forward tilt angle Φ defined for each frame as the angle between the vertical direction *y* and the axis binding the center of the hip and the head, expressed in radians (Figure 1.c). The resulting tilt angle sequence is described by the following 5 parameters: mean, standard deviation, ratio between the global minimum and maximum values, number of local maxima and relative temporal instant of the global maximum.

For the *Time* subcomponent of the *Effort* quality (8 features), we firstly consider the total gesture duration (number of frames). Then, we compute features based on the kinetic energy sequence defined for head and left and right hands. The frames of the sequence with kinetic energy inferior to 1/10 of the maximal energy reached throughout the gesture are considered as pauses, the others as high/medium activity frames. We compute the percentage of low activity frames relatively to the whole sequence, as well as the mean, standard deviation and maximum value of the kinetic energy sub-sequences for both high/medium and low activity components.

The *Flow* subcomponent of the *Effort* quality (10 features) is described with the help of the third order derivative of the left and right hands trajectories, so-called jerk. The 2 jerk series are statistically described by the following entities: mean, standard deviation, ratio between the maximal and mean values, number of local maxima and the relative temporal instant of the global maximum.

For the *Weight* subcomponent of *Effort* quality (30 features), we consider the vertical components of the velocity and acceleration sequences (*i.e.*, $y'_{.,t}$ and $y''_{.,t}$ signals) associated to 3 joints: the center of the hip, the left and the right hand. The following 5 features are retained: mean, standard deviation, maximal amplitude, number of local minima, and relative temporal instant of the global minimum value. Such an approach makes it possible to characterize the vertical motion of the gesture sequence.

Finally, for the *Shaping* sub-quality (24 features), a first part of the description aims at globally characterizing the spatial dissymmetry between the two hands over the whole sequence. For each frame, we associate a dissymmetry measure defined as: $Dys = \frac{d_{left,\,center}}{d_{left,\,center}+d_{right,\,center}}$, where $d_{left/right,\,center}$ denotes the distance between the left/right hand and the center of the shoulders. The Dys_t sequence is globally described by 6 parameters: mean, standard deviation, global maximum/mean ratio, global minimum/mean ratio, number of local extrema, and the relative temporal position of the global extremum. In addition, the same parameters are associated to 3 other sequences, respectively associated to global body amplitudes in the directions perpendicular to vertical, horizontal and sagittal planes (Figure 1.b), respectively denoted by A^x, A^y, and A^z and defined as $A_t^d = (max_i(|d_{i,t}^{trans}|) - min_i(|d_{i,t}^{trans}|))_{t=0}^{N-1}$, d ∈ {x; y; z} where *i* indexes the skeleton joints.

Table 1. Mean F1 scores (in %) obtained for the various classification method retained.

Gesture	SVC	Extra Trees
Iconic	98.8	99.5
Metaphoric	98.2	99.2
All gestures	97.2	98.7

Experimental results

We tested our expressive model on Microsoft Research Cambridge-12 (MSRC-12 Figure 2) dataset [7], which is publicly available and provides two different types of gesture captured by a Kinect camera: 6 gestures are *iconic* and represent actions/objects and 6 others are *metaphoric* (more related to abstract concepts). As proposed in [8], we have first performed the recognition separately on gestures of each given type (*iconic*, *metaphoric*). Then, we have considered globally all the 12 gesture categories, in order to test for scalability. For each of these 3 different recognition runs, two different methods were compared. The first one uses *Support Vector Classifiers* [9] with the *one versus one* strategy: there is one classifier for each pair of classes and at the testing stage, the class collecting the highest number of classifier votes, is retained. The second one is the *Extremely Randomized Trees* (Extra Trees [10]): at each tree node, a component of the feature vector and a threshold are randomly selected and split in left and right child nodes. When all leafs contain only one training sample the class is associated to and the process is stopped. At prediction time, the feature vector is processed by all the trees, and the class collecting the highest number of votes is retained. We use the F-score as performance measure: $F1\ score = 2\,\frac{precision\,.recall}{precision+recall}$. We have applied a 5-fold cross-validation scheme, *i.e.* with a training/testing ratio of 80%/20% and 5 cross-validation steps, by splitting the data into 5 blocks preserving the initial class distribution.

The classification results are summarized in Table 1. They correspond to the average F-scores obtained on all gesture classes involved (iconic, metaphoric, all). The results correspond to the best obtained among different parameters combinations associated with each classification strategy (SVC or Extra Trees) and obtained through an optimization strategy. The mean F-measures obtained are in all cases superior 97%, whatever the classification strategy involved. The testing of the proposed model on MSRC-12 dataset gives better results than any other approach having privileged the common use of joints positions or velocities sequences and having put aside the semantic interpretation of gesture in the designing of its descriptors [8] [11] [12]. This demonstrates the capacity of our descriptors to efficiently capture relevant motion indices for action recognition purpose.

Figure 2. Examples of motions tracked for MSRC-12 datasets constitution.

Conclusion and perspectives

In this paper, we introduced a gesture description approach, based on a set of descriptors dedicated to various entities defined in the LMA model. Our perspectives of future work concern the research of other descriptors still inspired by considerations on expressivity, their dimensional reduction to 2D context in order to test them on numerous available 2D video datasets and their application to other types of gestural contents.

References

[1] Chen, H.-S., Chen, H.-T., Chen, Y.-W., Lee, S.-Y. (2006). Human action recognition using star skeleton. In *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pp. 171-178. ACM Press.

[2] Kellokumpu, V., Zhao, G., Pietikäinen, M. (2008). Human activity recognition using a dynamic texture based method. In *BMVC*, pp. 1-10.

[3] Oikonomopoulos, A., Patras, I., Pantic, M. (2005). Spatiotemporal salient points for visual recognition of human actions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, ***36****(3)*, 710-719.

[4] Rapantzikos, K., Avrithis, Y., Kollias, S. (2009). Dense saliency-based spatiotemporal feature points for action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009)*, pp. 1454-1461. IEEE Press.

[5] Laban, R. (1994). *La Maîtrise du Mouvement*. Arles: Actes Sud.

[6] Zhao L., Badler, N.I. (2005). Acquiring and validating motion qualities from live limb gestures. *Graphical Models*, ***67****(1)*, 1-16.

[7] Fothergill, S., Mentis, H., Kohli, P., Nowozin, S. (2012). Instructing people for training gestural interactive systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pp. 1737-1746. ACM Press.

[8] Song, Y., Morency, L.P., Davis, R. (2013). Distributions-sensitive learning for imbalanced datasets. In *2013 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2013).* IEEE Press.

[9] Suykens, J.A.K., Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, ***9****(3)*, 293-300.

[10] Geurts, P., Ernst, D., Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, ***63****(1)*, 3-42.

[11] Negin, F., Özdemir, F., Akgül, C.B., Yüksel, K.A., Erçil, A. (2013). A decision forest based feature selection framework for action recognition from RGB-depth cameras. In *Image Analysis and Recognition*, pp. 648-657. New York, NY, USA: Springer.

[12] Jiang, X., Zhong, F., Peng, Q., Qin, X. (2014). Online robust action recognition based on a hierarchical model. *The Visual Computer*, ***30****(9)*, 1021-1033.

Articulated tracking of humans by video technology

Nico van der Aa[1,2], Coert van Gemeren[2], Remco Veltkamp[2], Lucas Noldus[1]

[1]Noldus Information Technology, Wageningen, The Netherlands
n.vanderaa@noldus.nl, l.noldus@noldus.nl
[2]Utrecht University, Utrecht, The Netherlands
c.j.vangemeren@uu.nl, r.c.veltkamp@uu.nl

Abstract

Measuring a person's behavior in any kind of domain by a computer system, starts with measuring and analyzing that person's movements. Video technology provides an unobtrusive way of capturing this information. In this paper, we will focus on articulated tracking, which is the field of estimating and tracking the body pose of a person. A pose consists of the joints and rigid parts of a person's skeleton at each frame. Alternatively, we use only the body orientation to estimate the direction at which a person is focused. We discuss the findings of our research on state-of-the-art methods in articulated tracking and body orientation estimation techniques with a distinct sensor setup to give researchers an idea of the challenges they might face for their application.

Introduction

To create tools for measuring human behavior in an automated way, the system must detect, track and analyze human motion. Camera technology provides a way to measure movements and activities in a scene unobtrusively. The field of capturing human motion is an important field within computer vision which is far from solved. A camera gives an array of pixels, including RGB colors for video cameras, or depth information for structured light cameras. The algorithm will provide details of the detected people as joint positions and orientations (e.g. of the torso). The main challenges are (1) to separate people in the foreground from the background by defining pixel regions which are unaffected by the endless amount of variability caused by clothing and background similarities, (2) to identify the limbs and cope with the symmetry of the kinematic structure of the human body, (3) to handle occlusions by objects, other people and self-occlusions, and (4) obtain these results in (near) real-time.

For articulated tracking, the literature divides these approaches in *model-based* (or *generative*) approaches versus *model-free* (or *discriminative*) approaches. Model-based approaches rely on an explicitly known parametric human model, and match the image observations to this predefined model. The kinematics of the model provide the basic restrictions for the human shape, such as the tree-structured kinematic constraints between adjacent body parts (e.g. torso-upper half-limb connection). In contrast, model-free approaches estimate a pose directly from observation, without using an accurate 3D model. They use the fact that the set of typical human poses is far smaller than the set of kinematically possible ones and train a model that directly recovers poses from observable image quantities. As our intention is to develop a tool to measure human motion in any kind of application, we restrict ourselves to model-based approaches. From literature, we select three model-based approaches based on a different sensor setup: (1) a single video camera; (2) a system of multiple calibrated cameras and (3) a depth sensor using structured light. The underlying idea is that the sensor and algorithm choice depends heavily on the application. As an example, a depth sensor has a range up to 8 meters and the infrared light principle will not work in an outdoor environment. In this paper we share our findings with the selected approaches.

For some applications, articulated tracking is not possible and also not needed. Think of a shopping center where many people walk around arbitrarily and we only want to know which direction they are facing or in which direction they are moving. Therefore, we have studied a state-of-the-art method for body orientation estimation.

Articulated tracking

In our research we consider three types of sensor setups to capture human motion: (1) a single static video camera, (2) a system of multiple static and calibrated video cameras, and (3) depth sensors using structured light. As we are interested in developing a general tool that is independent of the application, there are no constraints imposed on the pose to be detected. To keep it feasible, only a single person is assumed to be in the scene.

Monocular video camera

A single camera only provides a projection of the real world on a 2D image plane. The best strategy is to capture the projection of the skeleton on the camera view, which is the so-called 2D human pose. Pictorial structures is a model-based technique to estimate such 2D human poses. A pictorial structures model for a human being consists of a collection of body parts with connections between certain pairs of parts. It is a class of graphical models where the nodes of the graph represent body parts, and edges between parts encode pairwise geometric relationships. The appearance term (node) models the probability of a part being present at a particular location and orientation given the input image. A prior models the probability distribution over the pose, constraining the estimated pose to be plausible in terms of human articulation. To enable efficient inference, often two (unrealistic) assumptions are made: (i) the appearance of a part is assumed independent of its pose and that of the other parts; (ii) the prior over pose is a Gaussian with 'tree-structured' covariance. However, the main challenge for 2D pose estimation is the depth ambiguity, which occurs for example when an arm is behind the torso.

To estimate the 2D pose from videos, we analyzed the method of Ramanan *et al.* (2007), as it was tested on videos from various sources containing a wide range of activities and showing promising detection results at body part level accuracy. It employs pictorial structures in a tracking framework which includes a model-building phase and a detection phase. During the model-building phase the system selects a frame where the pose is distinctly present and all body parts are visible. This way it learns a model of the person who is to be tracked in the video sequence. In the detection phase, this model is applied to detect the current state of the pose in each frame of the video sequence. The main assumption of this method is that in both modules of the system, the scale of the person to be detected is known beforehand and it must remain approximately constant throughout the video. The system will not detect people wearing skirts, dresses or loose clothes, as body parts are modelled by rectangles. An example of the output is shown in Figure 1.

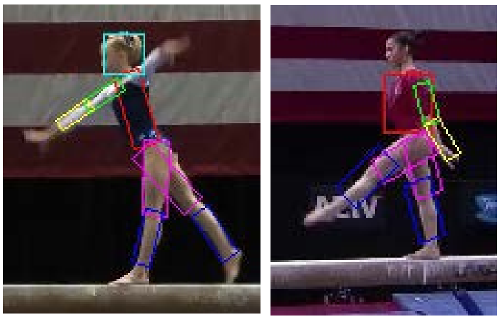


Figure 1. Examples of monocular pose estimation.

We analyzed the original method in terms of robustness and performance with respect to the type of input videos and found that, in contrast to what is stated in the original paper, the method is highly dependent on the set of input parameters, which do not translate across different videos. We also showed situations where one set of parameters can lead to different results within the same video, and we found that the motion model can eliminate the necessity to manually adjust these parameters when processing a single video. As a conclusion, the method under investigation does not solve the problem of occlusions as common by monocular pose estimation. The reader is referred to Ursu (2013) for more details.

Multiple static and calibrated cameras

Another way of handling occlusions is by including more camera views. Parts of the person that are occluded in one camera view, may still be visible in other views. If the cameras are static and calibrated, we know where they are with respect to each other and we can link their views to the real world.

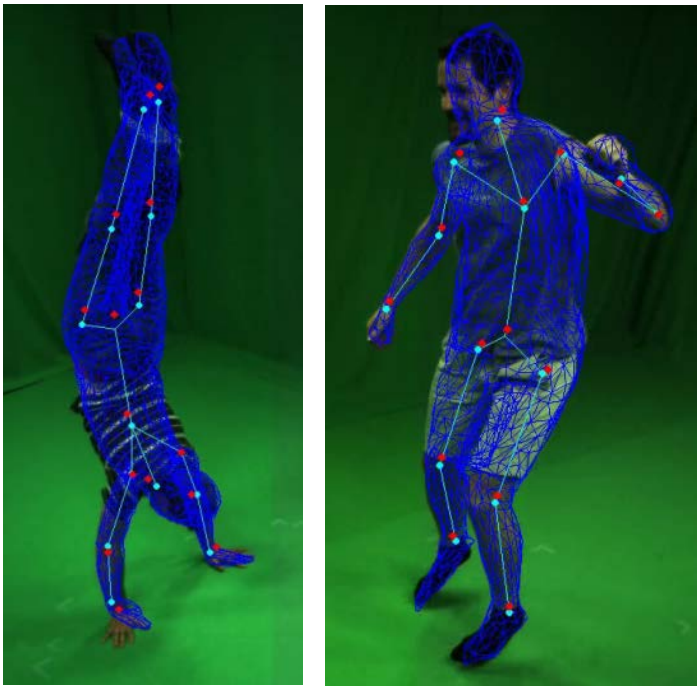


Figure 2. Examples of multi-view pose estimation. The white dots are our results and the red ones are the results of Gall *et al.* (2009).

We analyzed and extended the method presented by Gall *et al.* (2009), which is also a model-based pose estimation technique. The method estimates the pose by fitting a 3D model of the observed person to the images from the camera views. It accurately finds the skeleton pose and from this pose the mesh deformation is calculated by a three step algorithm. In the first two steps the motion of the skeleton and its shape are estimated. These steps consist of a combination of two optimization approaches: a local and a global optimization. For the local optimization a weighted least squares problem is solved to minimize the distance between the model and the silhouettes. The global optimization uses a combination of particle filters and simulated annealing to estimate the pose where the local optimization fails. Given the mesh model, camera calibration data and the camera video sequences, an accurate pose is estimated. The last step enhances the shape of the model to match the fine details of the observed person (like loose fitting clothes such as a robe or a skirt) in the images.

Our implementation of this method provides accurate results as shown in Figure 2, but fails in some cases for small body parts like hands and feet. Both the local and global optimization are based on silhouettes, so a proper background subtraction is required. The main challenge is the computational effort. Especially the global optimization step requires many iterations, causing the time taken to estimate the pose in each frame to be in the order of minutes. For a more detailed discussion on this method see Resodikromo (2012).

Depth sensor

With the introduction of structured light cameras like the Microsoft Kinect sensor, cheap alternative ways of capturing a scene have become available. Instead of 2D features or 3D matching, 3D depth information is now directly available. SDKs like NiTE (www.openni.org/files/nite/) include simple tools for body pose tracking. Although the applications are restricted to situations where the person is fully in the field of view and at most 8 meters away from the camera, NiTE provides a fast and relatively stable method for pose estimation. Similarly to the previous methods, the difficulties lie with the detection of the limbs as their appearance is the least distinctive and have the smallest dimensions, which result in wrong or lost detections of wrist and elbow joints.

In our research on 3D hand and finger tracking Koetzier (2014), we applied the 3Gear tool (www.threegear.com). Although it is necessary to put the camera within a meter from the hands, it is possible to combine the skeleton tracking with finger tracking. As the development of these SDKs will continue and successors of the depth cameras will follow rapidly, this provides a good starting point for applications within the requested constraints.

Body orientation

Instead of estimating the full body pose, which still faces many open issues, we can also restrict ourselves to finding the body orientation, which often is a sufficient cue for analyzing a person's activity including the focus of attention. Based on Chen and Odobez (2012), the body orientation of multiple human targets is estimated from a video sequence, captured by the view from a single moving camera. Accomplishing this goal requires a few stages, including human body detection and tracking. Additional computation, such as determining real-world 3D position coordinates of the targets, as well as its velocity and direction, can improve the results.

Estimating human body orientations can be formulated as a classification task with multiple classes of body angles. In the paper we discuss here, 8 angle classes are defined. We introduce a method that incorporates a set of different classifiers and cues, allowing us to be more flexible in choosing the classification methods, and to have the best results obtained from the combined response from several classifiers (committee). The input are video frames on which human tracking is performed using the method described in Choi *et al.* (2012), which is preferred over other tracking methods, as the input is allowed originate from a single moving camera. Besides, the method is able to provide estimates for the positions of the human targets in real world coordinates. This information is particularly useful to determine the velocity direction and magnitude, of the targets. Aside from the coordinates of the targets, the method also returns bounding boxes of the targets. From these bounding boxes Histogram of Gradient descriptors are extracted. These are then supplied to several pre-trained classifiers such as a Neural Network and a Support Vector Machine, which give probability estimates for each of the 8 angle classes.

Figure 3. Example of the body orientation method. Top green denotes the head orientation, bottom green the body orientation. The red lines denote the temporal smoothened directions of the head and body, respectively.

In our implementation (see Ichim *et al.* (2014)) we add the face as an important cue to the overall orientation estimation, since it restricts the plausible angles. To include this information, face detection is performed on the bounding boxes. To maintain the consistency of the probabilistic framework, a uniform distribution, based on the presence or absence of a face, is generated. Other information is gained from velocity direction and magnitude, which is integrated in the framework by fitting a standard Gaussian distribution, centered on the velocity direction of an angle class in such a way that a relatively high velocity yields a high probability for the frontal direction, and a low probability for the other directions. A relatively low velocity yields the same probability for all directions. The response from all of the mentioned classifiers and additional cues are combined, and the final angle estimation the one with the highest probability. However, the final result is filtered using a sliding window to ensure temporal smoothness of the change in orientation over time. Compared to Chen and Odobez (2012), we were able to reduce the average error by more than 20 degrees and reduce the computation time by 400 times making this method near real-time. An example of the output is given in Figure 3.

Concluding remark

Human pose estimation from video technology remains a challenging field in computer vision due to occlusions in general, the dimensions of the different body parts and their variability in appearance. Obtaining human body orientation is feasible in real-time.

Acknowledgement

This research was supported by the FES project COMMIT/ and the NeurAS research project, which is part of IMDI-NeuroControl.

References

[1] Chen, C., Odobez, J. (2012). We are not contortionists: coupled adaptive learning for head and body orientation estimation in surveillance video. *IEEE Conference on Computer Vision and Pattern Recognition 2012.*

[2] Choi, W., Pantofaru, C., Savarese, S. (2012). A general framework for tracking multiple people from a moving camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence, **35**(7),* 1577-1591.

[3] Gall, J., Stoll, C., Aguiar, E.d., Theobalt, C., Rosenbahn, B., Seidel, H.-P. (2009). Motion capture using joint skeleton tracking and surface estimation. *IEEE Conference on Computer Vision and Pattern Recognition.*

[4] Ichim, M., Tan, R., Aa, N. van der, Veltkamp, R. (2014). Human body orientation estimation using a committee based approach. *9th International Conference on Computer Vision Theory and Applications.* Lisbon.

[5] Koetzier, M. (2014). *3D hand tracking using depth sensors.* Master thesis, Utrecht University, Information and Computing Science, Faculty of Science. Retrieved 22-6-2014, from < http://dspace.library.uu.nl/handle/1874/291559>

[6] Ramanan, D., Forsyth, D., Zisserman, A. (2007). Tracking people by learning their appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, **29***(1), 65-81.

[7] Resodikromo, J. (2012). *Marker-less 3D pose estimation.* Master thesis, Utrecht University, Information and Computing Sciences, Faculty of Sciences. Retrieved 22-6-2014, from <http://dspace.library.uu.nl/handle/1874/255396>

[8] Ursu, E. (2013). *Pose Estimation in Video.* Master thesis, Utrecht University, Information and Computing Sciences, Faculty of Science. Retrieved 22-6-2014, from 003Chttp://dspace.library.uu.nl/handle/1874/282663>

Observing Human Behaviour to Identify Risk in Task Performance

Ronald Poppe and Mark ter Maat

Information and Computing Sciences, University of Utrecht
Human Media Interaction, University of Twente
r.w.poppe@uu.nl

Introduction

Small mistakes in task performance can have large consequences. Administering a wrong dosage of a medicine, ignoring a system warning or failing to check whether certain values are set can have severe negative effects for patient and process safety. The repetitive nature of such tasks and the often high cognitive load of the worker can easily lead to decreased attention, with inaccurate or incorrect task performance as a result. These effects are even stronger when the normal workflow is interrupted, for example when the worker receives a phone call or when other unexpected events occur. These deviations from a typical task performance can be indicative of risky task performance. By observing the behaviour of a worker, we aim at identifying these moments in time and to alert the worker. We specifically focus on training scenarios in which we can provide understandable feedback to the worker, so the risky behaviour can be avoided in the future. In this paper, we present the Patria system, which observes gestures, head movement, facial expressions, eye gaze and key strokes from subjects operating a simple medical device.

User Observation

The Patria system's recording setup consists of several visible-light and infra-red cameras, infra-red lights, a 19" desktop monitor and computers to process and log all input and to run the simulation software (or any other software that the worker needs to work with). Figure 1 shows an overview of the setup.

Two visible-light cameras below the screen are used to analyse facial expressions and head orientation, whereas the two infra-red cameras are employed for gaze estimation. The two pairs of cameras are calibrated and synchronized to allow for stereo-images to be recorded. Especially for the estimation of gaze, the distance from the camera to the eye is of great importance for robust analysis. The two cameras above the screen (see Figure 1) are used to analyse the hand gestures of the worker. The placement of the cameras is such that the worker can freely move around in a relatively large space.

Figure 1. Overview of the Patria system.

For the processing of these image streams, several software programs are employed. For the gaze estimation, we rely on the ITU gaze tracker (San Agustin et al., 2010), which proved to be reasonably robust in the detection of whether a person looks at the screen or not. We have not attempted more accurate detection and consider this a viable next step in the further development of the setup.

The analysis of facial expressions and 6 degrees of freedom head pose (location and rotation) is performed with VicarVision FaceReader 5.0 (Den Uyl & Van Kuilenburg, 2005). Gestures are analysed by fitting a parametric body model to the stereo-images obtained from the top cameras. Finally, all inputs, including videos and estimated behaviour parameters, are time-stamped and logged in Noldus Observer XT 11.0 (Zimmerman et al., 2009), and can be replayed and further analysed when desired.

Workflow Monitoring and Risk Assessment

One important aspect of the Patria system is its ability to estimate how risky the observed behaviour of the worker is. For this, we employ two different mechanisms:

Work-flow monitoring: for the analysis of risk, we take into account the notion of a correctly performed task. Many workflows, notably those that can have severe consequences for safety, have strict regulations to minimise these risks. Deviating from this normal task performance is considered risky behaviour. To this end, we model a task flow as a state diagram. Each state in the diagram corresponds to a state in the workflow. Transitions between states are triggered by detected behaviour events such as looking at the screen, pressing “enter” or picking up a form with the left hand. Transitions can also occur based on a timer. For example, if a certain action does not occur within 5 seconds after arriving in a certain state. Each transition has an associated risk value assigned to it, between 0 and 100%. As both the states and the actions triggering the transitions between them are interpretable, we can present specific information to the worker about why the risk value was increased. This can help in understanding and improving the task performance, both in training situations as well as in online use.

User state monitoring: based on the observed facial expressions, described as action unit activation scores, we try to find out whether or not they correlate with these moments in time where risky behaviour occurs. Intuitively, one might expect that a stressed, bored or frustrated person is more likely to make mistakes. Also, we expected that people would show observable facial reactions to unexpected events, which are also more likely to result in abnormal task performance. To estimate the level of risk from a worker’s facial expressions, we have trained a support vector machine (SVM) classifier based on pre-recorded data in which we manually identified risky behaviour. The SVM classifier is then evaluated on a frame-by-frame basis to arrive at a user state estimate over time.

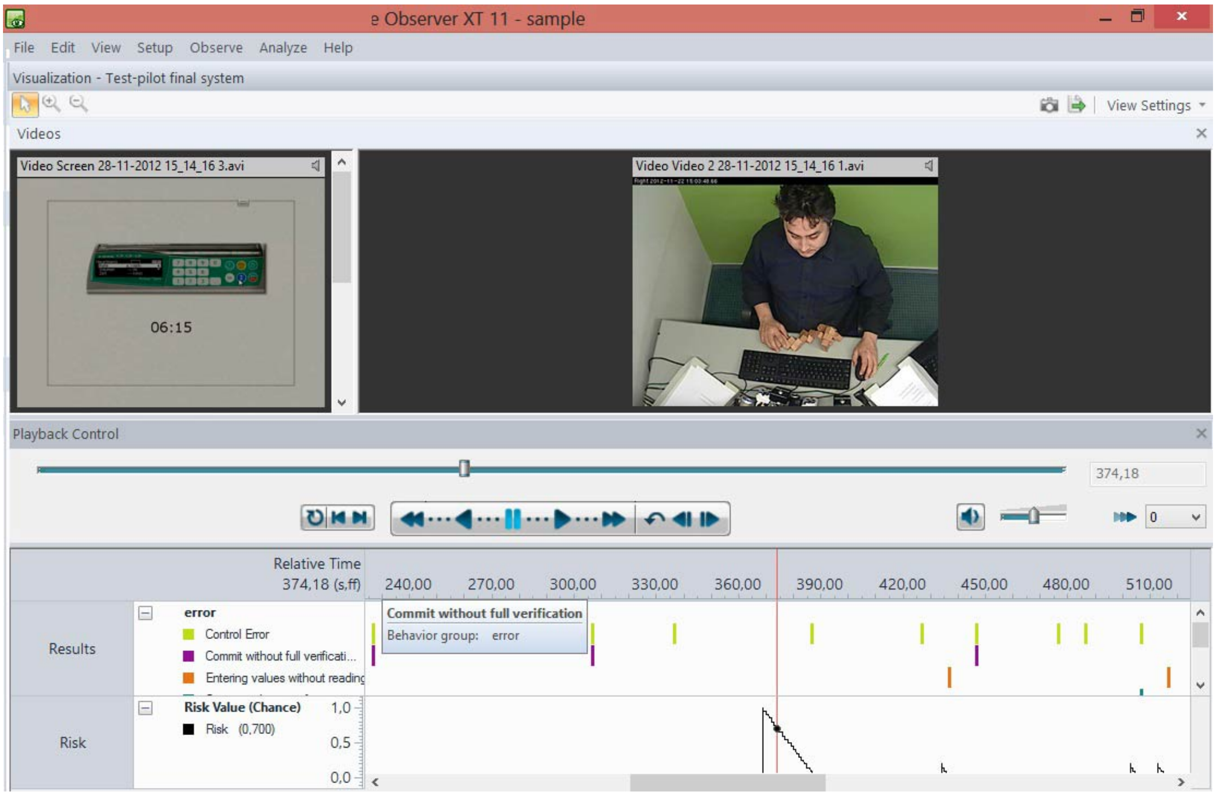


Figure 2. Screenshot of the risk assessment module in Noldus Observer XT 11.0.

Risk values are summed over time, and small risk values added in rapid succession lead to an increasing risk value over time. A decay function is used to lower the risk after each frame. Apart from the analysis of these values over a task performance, a threshold is set on this risk value to allow for the trigger of alerts to the worker or a supervisor. See Figure 2 for a screenshot of the risk estimation module in Noldus Observer XT 11.0.

Discussion and Future Work

We have applied the Patria system in a training scenario where subjects had to set the values of an infusion pump (see Figure 2, left-upper corner) based on patient information. We considered three different executions of the task: (1) read the correct value from a screen and enter it, (2) read the correct value from a form and enter it, (3) read the correct value from the screen, fill it in on a form and enter it in the simulator. This workflow was modelled in a state diagram. Risky actions were assigned when the subject did not check the number that was filled in, when the subject failed to finish the task in time, when the form was not taken or stored at the right location, or when several of the actions were repeated within one session. At semi-random moments, the subject was prompted with a beep to perform one of the three tasks. In the meantime, to increase the cognitive load of the participants and to ensure that enough non-task-related behaviour was displayed, the subjects were to solve a wooden snake puzzle. Currently, we make risk decisions mainly at the end of a "turn". However, our rules could also have a temporal component, such as action B should be performed within 10 seconds after action A.

We tested the system using 20 participants. Overall, risk values corresponded with episodes in which the subjects made mistakes, usually by not checking the entered values, or by not finishing before the next beep. In some cases, a missing detection of gaze towards the screen resulted in an estimated risk of not having verified what was entered on screen. More robust eye gaze estimation could solve this issue. We found that the use of facial expressions did not add to the performance of the system, typically because subjects were often not showing much variation in their facial behaviour. We expect that the monitoring of user state from the face becomes more important and informative when the task performance is measured over extended periods of time. Boredom and frustration then might be more salient.

In the future, we plan to reduce the number of components in the system, notably the number of processing units. Also, we will combine most of the equipment in a single device. Eventually, we aim at designing a robust low-cost worker observation system that is portable and can be adjusted to different workplaces. From a processing point of view, we will address automatically learning the state diagrams for normal and abnormal task performance, and facilitate the process of assigning risk values to state transitions. Especially when several tasks are performed in parallel, this is a challenging process.

Acknowledgements

This work is supported by the INTERREG IV A project Patria2 (http://patria2.eu/), and partly financed with the European Fund for Regional Development (EFRO), coordinated by Euregio. The authors like to thank the partners for their collaboration in the project: Noldus (Tobias Heffelaar), VicarVision (Marten den Uyl, Jordi Bieger, Mark de Greef), Fachhochschule Münster (Uvo Hölscher, Michael Lindenthal, Heiko Verwold).

References

[1] Den Uyl, M.J., Van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. In *Proceedings of the International Conference on Methods and Techniques in Behavioural Research (Measuring Behaviour)*, Wageningen, The Netherlands, pp. 589-590.

[2] San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D.W., and Hansen, J.P. (2010). Evaluation of a low-cost open-source gaze tracker. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA)*. Austin, TX, pp. 77-80.

[3] Zimmerman, P.H., Bolhuis, J.E., Willemsen, A., Meyer, E.S., Noldus, L.P.J.J. (2009). The Observer XT: A tool for the integration and synchronization of multimodal signals. *Behavior Research Methods, **41**(3)*, 731-735.

A multimodal benchmark tool for automated eating behaviour recognition

V.D. Kakra[1], N.P. van der Aa[1], L.P.J.J. Noldus[1] and O. Amft[2]

[1]Noldus Information Technology, Wageningen, The Netherlands
v.kakra@noldus.nl
[2]Department of Sensor Technology, University of Passau, Passau, Germany

Abstract

In this paper, we present the multimodal eating behaviour dataset called iEatSet (***i***CareNet Multimodal ***Eat***ing Behaviour Data***Set***). iEatSet shall serve as an algorithm benchmark dataset and aims to facilitate research in automatic dietary monitoring, eating recognition, and activity recognition in general. iEatSet provides multimodal synchronised data streams, including multi-camera vision, inertial motion sensor data, and associated ground truth labelling, recorded from 15 participants over 5 meals in a natural restaurant environment. Recordings included food selection and consumption without scripted protocol, to provide naturalistic behaviour.

Having validated methods and tools to recognize people's eating behaviour could advance research and coaching in applications related to nutrition and dieting. Currently, dieting analysis is done manually and is very tedious. Hence, an automated analysis tool is desired. The current state-of-the-art tools in activity recognition first need to be trained before they can recognize activities. The iEatSet can be particularly useful to benchmark supervised recognition algorithms and serve as reference for unsupervised algorithm analysis.

Introduction

Significant research has been performed in the field of human eating behaviour that includes different aspects like the monitoring of food intake, food selection choices and eating habits of a person. We focus on recognition of a person's food intake activities like the number of bites or sip taken. The counting of number of sips and bites is vital to interpret eating behaviour which can be used to coach the patients and counteract eating disorders that include obesity and anorexia nervosa.

Currently, studies that analyse food intake activities have human observers who use tools like The Observer XT (www.noldus.com /observer), a software package for the collection, analysis and presentation of observational data. In other studies, participants are requested to manually note down details about his/her food consumption. This analysis can be made more efficient if the recognition procedure is automated and standardized.

There have been different approaches developed and sensors used by researchers to automate the recognition of food intake activities as part of daily life activities. For example, Cheng et al. (2011) recorded senior citizens in their homes with a colour camera to recognize different eating activities and walking activities. These video recordings are used to assist in monitoring the health and well-being of the senior citizens. Amft et al. (2005) monitored a person's diet with the help of on-body inertial sensors to recognize eating behaviour for food intake. Stein et al. (2009) used a multimodal approach to recognize food preparation activities. They used inertial sensors along with cameras to detect activities, which improved the activity recognition results. Most of these existing methods for activity recognition use a machine learning approach, where they first train the classifier on training data and subsequently validate the method on test data. Both, training and evaluation of algorithms can benefit from a publicly available benchmark dataset for automatic eating activity recognition.

There are several activity recognition datasets available that use just video cameras or a multimodal sensor setup and some of them also record eating activities among other activities. For example, the Hollywood2 dataset by Marszałek et al. (2009) is a vision-based dataset containing 'eating' actions among 12 actions collected from different Hollywood movies, while the Senior Activity Recognition Dataset (SAR) by Cheng et al. (2011) is in the real world focusing mainly on recognition of eating classes and walking classes of senior persons. These datasets focus on detecting different activities rather than recognizing different eating activities. There are also multimodal datasets available for activity recognition by sensor fusion. One such multimodal dataset, which is closely related to what we want to record, is the 50 Salads Dataset by Stein et al. (2013) that records people preparing food. This dataset uses the Kinect v2 sensor (http://www.microsoft.com/en-us/kinectforwindows/) to record RGB-D data from a top view and has accelerometers attached to the objects like the knife, mixing spoon and other objects used for preparing the food. We use the same sensors but our focus is on food intake activity recognition and not on food preparation. The food intake activities involve movements of a hand towards the mouth, which is best captured from the front view with the upper body of the person completely visible. In this dataset, recording is done from the top view with only the hands of the participants visible so that is why the 50 Salads Dataset cannot be used and we recorded the iEatSet for the food intake activity recognition.

The iEatSet provides a dataset with 15 participants, recorded 5 times having different meals for 5 days and ground truth labelling is provided to use for both testing and validation of future algorithm development for eating behaviour as well as other activity recognition algorithms. The dataset includes uncompressed data from 3 RGB cameras that record the person from the front, side and top views, the calibrated data from the 4 inertial sensors that are attached to both the upper and lower arms and the data from the Kinect v2 depth sensor from the front view. All the data is time synchronized and the time synchronization is also provided.

iEatSet (***i*****CareNet** ***Eat*****ing Behaviour Data*****Set*****)**

Choice of Sensors

Food intake activities include actions like taking a bite or sip that involves movement of the hand from the plate/table towards the mouth. This movement can best be captured through vision by capturing the upper body of the person with the hand and face of the person visible. That is why we recorded the RGB data from the front view. Also, the distance between either the food or cutlery or food containers and the mouth can be used to recognize whether the person is eating or not. This information can be captured with a depth sensor. The top view cannot be used here because the person's mouth isn't visible and if the person leans over to consume food, only the head of the person would be visible while most other things would be occluded. That's why we choose to use the Kinect v2 to capture the RGB-D data from the front view. An additional Kinect v2 sensor cannot be used to capture the depth data from the side or top view because this would cause interference while recording.

One big disadvantage of using cameras is that there can be occlusion at any given time during the recording. This occlusion problem can be tackled by recording from different field of views using multiple cameras. The person can be captured eating from the side view and the top view. The movement of hands towards the mouth can be captured best from the left side for a right-handed person and on the right side for a left-handed person. This way the camera can get a clear view of the hands, cutlery and the table and the eating activities can be recognized from the side view. The top view can be used to observe the food items on the plate and the table as well as the movement of the hands when a person eats. This information can be used to recognize food intake activities. The top view also provides researchers a way to investigate food identification. That is why we used 2 IP cameras to capture RGB data from the side view and top view that includes the table and participant while eating. Each of the three camera views can be used separately to recognize eating activities or a multi-camera view approach can be used.

Figure 1. Frontal view of person eating.

On-body sensors like the inertial sensors can also be used to recognize the eating activities. There has already been research in this area where the Inertial Measurement Units (IMUs) or inertial sensors have been used to recognize eating activities. Amft *et al.* (2005) demonstrated inertial body-worn sensors can be used to detect eating and drinking gestures from other gestures where the sensors were placed on both the lower and upper arms of the person. When a person eats, he/she holds the food or cutlery in a specific way which is reflected in a specific orientation of the hand which can best be captured by a gyroscope. Then the person moves hands with the food towards the face and the speed of the hand can be captured by the accelerometer. These features can then be used to recognize eating activities. The orientation and acceleration can be captured when the sensors are placed on the wrists or lower arms and upper arms of the participant.

Sensors Description

Kinect v2: The Kinect v2 records the RGB data with a resolution of 1920×1080 pixels @30fps and the depth sensor with a resolution of 512×424 pixels, 13-bit depth. The RGB data and depth map are synchronized internally by the Kinect v2. The Kinect v2 recorded the person eating from the front view as shown in Figure 1.

RGB camera: Axis M1054 IP cameras were used to record the person from the side and top views with a resolution of 1280×800 pixels @25fps. The calibration between all the three cameras was done by using a checkerboard pattern and the intrinsic and extrinsic parameters are made available. One camera was placed above the table and little in front of the person so that the plate was visible at all times even if the person leans forward to eat. The other camera was placed on the left side if the person was right-handed and on right side if the person was left-handed.

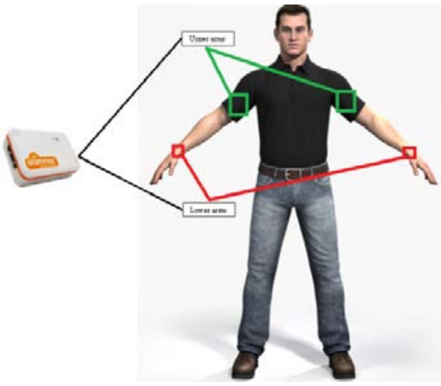


Figure 2. Shimmer3 sensor placement.

Inertial Sensors: Shimmer3 (www.shimmersensing.com/shop/shimmer3) is a state-of-the-art wearable IMU sensor which has an in-built accelerometer, gyroscope and magnetometer. The size of Shimmer3 is similar to that of a wrist watch so it can easily be mounted. Data can be recorded on the SD card slotted on the Shimmer3 sensor or streamed directly to the machine via Bluetooth. Data was recorded on the SD card of each Shimmer3 sensor. The sensors were mounted on the upper and lower arms of the person. The Shimmer data is 48-bit data from accelerometer, gyroscope and magnetometer @51.2Hz. The 4 devices are synchronized using the synchronization software (Multi Shimmer Sync for SD v1.2) where a master-slave pairing is established between the sensors. The Shimmer3 sensor is shown in Figure 2.

Sensor Network Synchronization

For synchronization between all sensors, the participant was asked to clap his hands twice: before the start and at the end of eating meals. This was because the clap is clearly visible through vision (Kinect v2 & RGB camera) and that would be used as a reference to start and end the recordings. A clap represents a sudden movement and acceleration of the hands which is clearly visible in the acceleration data from the Shimmer3 sensors. The synchronization is achieved using the timestamps generated by each sensor.

Study Design

The iEatSet consists of 15 participants where each participant was recorded when he or she was having lunch for 5 days. We expected 8-10 eating instances and 6-8 drinking instances per meal from each person. That was why we chose to record the same person for 5 days to have enough eating and drinking instances for training and testing for a single person's behaviour recognition. We chose to record 15 people because this gave us enough instances for training and test data for inter-person recognition. Each day the participant was served different food like sandwiches, hot meal, soup and yoghurt and different types of drinks like soft drink, orange juice, tea and coffee to gather as many instances and variety of eating gestures. The participant was asked to eat either with their hands or cutlery depending on the food served to ensure all the types of eating gestures were covered. Also, different drinks were served in a glass, bottle or cup, to ensure each day different types of drinking activities were recorded.

Eating gesture lexicon & ground truth annotation

The eating activities were annotated according to the classes provided by Koenderink *et al.* (2014). The annotation is two-layered: the first layer broadly classifies all the eating activities into 6 main classes and the second layer consists of 22 eating activities that are further grouped under each of these 6 classes. We modified this lexicon where we didn't annotate the gestures that happen inside the body and couldn't be detected by the sensors. With the two layers, researchers can investigate methods either for inter-class or intra-class eating activity recognition or a combination of them for both layers. The 6 main classes and the 22 sub-classes classified under these 6 classes are mentioned in Table 1.

Manual annotation is subject to human errors which could lead to the annotation being diverged from what it should be in multiple ways as described by Mathet *et al.* (2012). Each manually annotated element can diverge either by being annotated in the wrong location, the category to which it is annotated is not correct, is a false positive or a false negative annotation. In order to keep these errors to a minimum, we used a methodology whereby all experts label a share of the data and 20% in parallel. For the parallel annotated data, we provide iterator reliability metrics and timing coherence metrics. The Observer XT is used as the event logging software.

Table 1. The two layers of the eating gesture lexicon to be used in the annotation.

Layer 1	Non-eating gestures	Prepare food gestures	Prepare consumption gestures	Consume gestures	Finalize consumption gestures	Other
Layer 2	Play (moving food, plate); Wipe clothes, hand, mouth or table; Pick up cleaning aid; Stack waste or table-ware	Cut; Open package; Pick up package; Place spread on bread; Pour; Put down package; Stir;	Pick up cutlery; Move to mouth; Pick up food or drink;	Take bite; Take sip;	Move away from mouth; Put down cutlery; Put down food;	Lick finger, cutlery or lips; Pick teeth; Express with hands;

Discussion

This dataset is made available to the research community. It contains (1) uncompressed and synchronised RGB videos of the recordings from the IP cameras and the Kinect, (2) the calibration data of each camera including the intrinsic and extrinsic parameters, (3) 13-bit depth data from the Kinect, (4) the raw, calibrated and synchronised 48-bit data from all the 4 IMUs, (5) the labelled annotations, (6) the timestamps generated by all the sensors and (7) accompanying software to read the data. With this benchmark, we hope to advance research in eating behaviour recognition and activity recognition in general.

Acknowledgement

This project was supported by the EU FP7 Marie Curie Initial Training Network iCareNet under grant number 264738.

References

[1] Amft, O., Junker, H., Troster, G. (2005). Detection of eating and drinking arm gestures using inertial body-worn sensors. In *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*, pp. 160-163.

[2] Cheng, H., Liu, Z. Zhao, Y., Ye, G. (2011). Real world activity summary for senior home monitoring. 2011 IEEE International Conference on Multimedia & Expo. (ICME), pp.1-4.

[3] Koenderink, N.J.J.P., Top, J.L., van Doorn, A. (2014). Unravelling the language of eating. *Workshop on Smart Technology for Cooking and Eating Activities (CEA2014).*

[4] Marszałek, M., Laptev, I., Schmid, C. (2009). Actions in context. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2929-2936.

[5] Mathet, Y., Widlöcher, A., Fort, K., François, C., Galibert, O., Grouin, C., Zweigenbaum, P. (2012). Manual corpus annotation: Giving meaning to the evaluation metrics. In *Proceedings of the International Conference on Computational Linguistics (COLING 2012)*, pp. 809-818.

[6] Stein, S., McKenna, S.J. (2013). Combining Embedded Accelerometers with Computer Vision for Recognizing Food Preparation Activities. In Proceedings of *the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013)*, Zurich, Switzerland.

Video Surveillance for Behaviour Monitoring in Home Health Care

Ben J.A. Kröse[1,2], Tim van Oosterhout[1], Gwenn Englebienne[2]

[1]Digital Life Centre, Amsterdam University of Applied Science, Amsterdam, the Netherlands
{B.J.A.Krose,T.J.M.van.Oosterhout}@hva.nl
[2]Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
G.Englebienne@uva.nl

Introduction

As the number of elderly needing assistance in their daily life increases, sensing systems that monitor behaviour are becoming ever more important. The type of the behaviors that are to be monitored varies enormously, from nutritional intake, physical activities, cognitive activities, fall detection, wandering detection or sleep behaviour. Also the types of sensing systems vary accordingly. Wearable sensors are often used for measuring physical activities and for alarming. However, these sensors require special attention from the user, in terms of battery replacement and compliance. Ambient sensors are more reliable in that sense, but result in another trade-off. Simple sensors such as motion detectors and pressure mats are non-obtrusive but are less accurate in their measurements. Advanced imaging sensors give much more accurate data but come with the problem of privacy. In the Balance-IT project, we specifically studied the applicability of imaging sensors in health care applications, specifically fall detection, wandering detection, remote physical therapy and fitness.

Falls are a major source of injury for elderly people (Gallagher et al., 2001). A lot of research has gone into automated solutions. Various fall detection methods using standard cameras have been presented, where rapid changes in the modelled body shape (Foroughi et al., 2009;Yu et al., 2009) or estimated pose (Liu et al., 2010) or bounding volume (Tao et al., 2005; Anderson et al., 2009) are used to detect falls. A major issue with these approaches is that how much the perceived body shape changes depend both on the camera location and the direction of the fall. This is not an issue in depth cameras, since they measure the absolute size of the body shape, and not the size of the projection of the body onto a 2D plane. Several methods have investigated the use of depth cameras, the Kinect in particular (Rougier et al., 2011; Mastorakis and Makris, 2012). Yet the body pose estimation of 3D sensors may fail dramatically in particular situations, such as in bright sunlight or when the person is partially occluded. In (Josemans et al., 2013), we investigated how combining 2D and 3D environmental cameras improves fall detection.

Wandering detection is a problem for people with dementia. Especially in nursing homes there is a need for automatic detection of wandering behaviour. In the Balance-IT project we focus on camera-based methods for wandering detection. Similarly to fall detection, it requires the precise detection of the patient's location in the environment, but it cannot be detected from short-term features: instead, longer-term tracking of the person is required and the tracks themselves are indicative of wandering. In (Martino-Salzman et al., 1991), wandering is distinguished from efficient ("direct") travel based on the path the patient takes, and is further subdivided into random motion, pacing (walking back and forth between two points) and lapping (walking in circles). (Vuong et al., 2011) directly implemented a rule-based classification algorithm to detect the 4 types of motion ("direct", "pacing", "lapping" or "random") and obtained promising results on a limited data set.

Although the current scientific trend is focused on using computer vision for fall- and wandering detection, such systems are not yet deployed at large scale. Probably a combination of costs, non-robustness and privacy issues play a role. Recent commercial applications also show a trend towards the use of the smartphone for this (Igual et al, 2013). In this paper we present two of the four Balance-IT projects that focused on computer vision: fall detection by RGB-D cameras and wandering detection with off-the-shelf video cameras.

Fall Detection

In practical applications, computer vision systems need to deal with the cameras' limited field of view, changing light conditions, occlusions, and varying projections. Active 3D sensors (RGB-D cameras) such as the Microsoft Kinect, provide a partial solution to both changing light conditions (by being active) and varying projections (by providing depth information), but are still subject to the sensor's limited field of view and to random occlusions. By combining multiple cameras looking at the scene from a different vantage point, we can improve on any single camera's detection results. In particular, the Kinect's pose estimation can fail quite dramatically when the subject is occluded, or when the subject's pose is unusual (e.g., lying down). In (Josemans et al., 2013), we investigated how the data from skeleton tracking could be fused with the input from a conventional camera for improved fall detection.

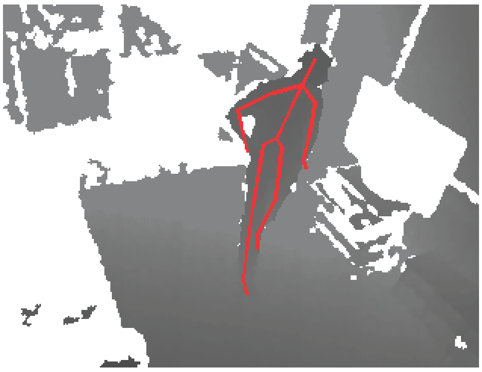


Figure 1. Skeleton as derived from Kinect depth image.

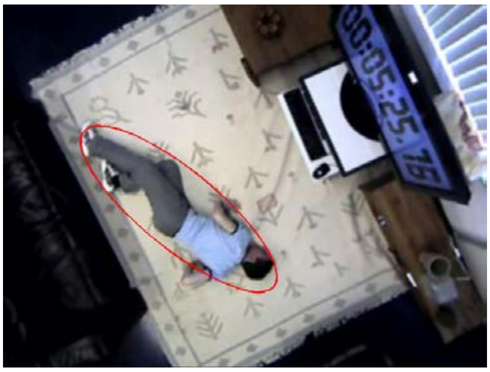


Figure 2. Bounding ellips extracted from foreground segmentation.

Our approach is based on combining data from a Kinect camera and a wide angle camera mounted on the ceiling. For the Kinect system we used the skeleton extracted from the depth image, see Figure 1. The 60D vector (20 joints) is reduced by a Principal Component Analysis (PCA). For the overhead image we fit an ellipse to the detected foreground from the conventional camera, as illustrated in Figure 2, and create a feature vector that includes the ratio between the ellipse's width and height and its location within the camera's field of view. First we tested a naïve fusion technique by just concatenating the two feature vectors. The results of the individual and combined features are given in Table 1. As can be seen this naive fusion approach does not yield improved results over individual features. Our approach, shown in Figure 3, validates the Kinect output with the conventional camera. To do this, we re-project the skeleton found by the Kinect into conventional camera's image. We then compute the overlap between the detected foreground in that camera and the projected skeleton, resulting in a "Skeleton Match Score", and feed this as an additional feature to our classifier. By using a non-linear Support-Vector Machine with RBF kernel, we obtain a classifier that correctly weighs the most informative features, i.e. the skeleton features when the skeleton is validated, and the conventional camera's features when it is not.

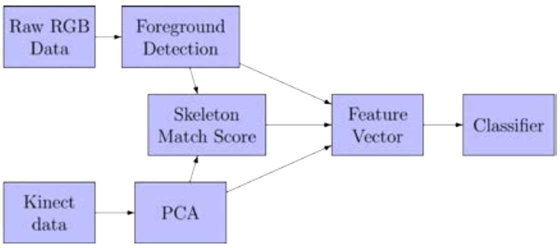


Figure 3. approach showing the skeleton match score.

Table 1. Performance of the four methods. TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negatives.

	TP	TN	FP	FN
Skeleton based	38.76	37.04	2.96	1.24
Bounding ellipse	36.28	40.00	0.00	3.72
Using all features	36.84	36.80	3.20	3.16
Using also skeleton match score	39.36	39.48	0.52	0.64

We recorded falls of students in our lab. The recorded falls all take place in an area of roughly 3x3 meters, and the direction of the fall is varied. In total, 40 fragments of about 5 minutes were recorded. Each fragment contains mostly general activity with a fall at the end. These falls were manually annotated.

Our results, shown in Table 1, with 4-fold cross-validation on 25 runs using random permutations of a dataset containing 40 falls and 40 non-fall sequences show that this approach substantially outperforms our baselines: the two methods based on a single sensor and the naive fusion method. In fact, only one fall sequence is not well recognized. In this sequence, the subject falls largely outside both cameras' field of view, so that both the skeleton-based and the conventional camera-based features are misleading.

Wandering detection

A second application area in the Balance-IT project is wandering detection. In contrast to fall detection, which focuses on an accurate estimate of the human pose, here we focus on an accurate estimate of the track: firstly to distinguish people with dementia from other people in the home, secondly to raise an alarm if the person with dementia is walking in an unwanted area (near staircase or other persons' apartments). We focused on a single camera.

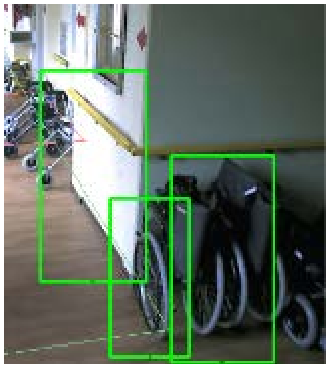


Figure 4. False positives in using HOG features.

We implemented person detection with HOG (Histogram of Oriented Gradients) features, which has been successfully applied to pedestrian detection. We found that most methods suffer from false positive detections when presented with a cluttered environment. To solve that, we developed a method that uses knowledge about the spatial structure of the environment. Many false positives appear in the wrong location, e.g. inside walls, or at the wrong scale, e.g. too small for the surroundings or with the feet not touching the ground (Figure 4). Furthermore, the range of searched scales must be limited in order to limit the computational complexity, resulting in missed detections either very close to or far away from the camera.

For an overview of the wandering detection see Figure 5. Using well established camera calibration techniques we obtain parameters for the camera's lens and position. In addition we define one or more areas on the ground that combined become our region of interest. Only this area is searched for the presence of people.

Traditionally a Support Vector Machine (SVM) classifier is trained on the HOG feature vectors of upright people. Then, an image is transformed into HOG features after which the trained feature vector, a window of fixed size, is swept across the transformed image. The process is repeated for transformations of the image at different scales. The result consists of detections across the whole image at different scales. Some position/scale combinations could have been eliminated beforehand considering the expected size of people and other constraints implied by perspective distortion and other environmental knowledge. We improve on this process by limiting the search area and scales using the manually defined region of interest and camera calibration.

For a certain distance to the camera an average person height can be computed in image space. Equidistant points form horizontal lines in the image. Therefore, an area containing several people standing at the same distance from the camera can be contained in a rectangle. If we determine a set of these rectangles, the widths, horizontal and vertical positions of which shall be determined by our region of interest and the height by person height and vertical position, we obtain a constrained set of search locations, all of which will result in people at the correct size and expected locations considering the information contained in the camera calibration and region of interest definition. If we run the HOG feature detector at a fixed scaled determined by the rectangle height for each such rectangle instead of at different scales across the entire image, the result is not only a higher true positive rate but also a lower computation time.

The detections from the method are tracked over time by a single hypothesis multi target tracker which uses Kalman filters for position extrapolation and smoothing. These tracks are classified by a rule based system on the basis of duration, covered distance and maximum speed into a walking or wandering class (pacing, lapping and random walks will all result in a change of these properties). Furthermore, regions within the region of interest determine which areas are safe and for which areas an alert should be generated. With this addition an alert can be generated for a wandering pattern that moves towards an alert region even before the person reaches it (Figure 6). Additionally alerts can be omitted if necessary for tracks originating from an alert region and moving away from it. We compared our method to a traditional multiscale HOG feature based people detector as implemented in OpenCV 2.4.8. On the same hardware our method ran 7 to 8 times faster while generating less false positives and less false negatives. Our method was not susceptible to certain patterns in the background generating consistent false positives in the compared method, while being more sensitive to people appearing at smaller scales because the search space could include finer scale differences at further distances. Finally, because of these more reliable detections our method caused less split tracks and less tracks consisting entirely or for a significant portion of false detections.

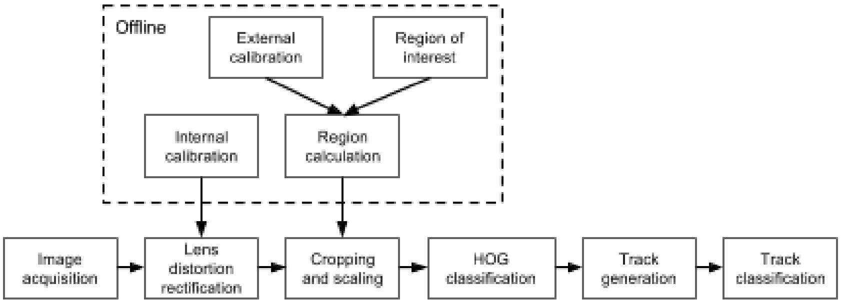


Figure 5. Set-up for the wandering classification.

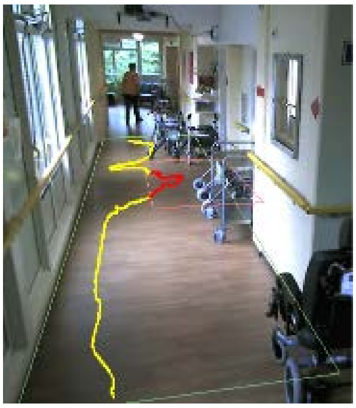


Figure 6. Results of our method.

Conclusions and discussion

Automatic monitoring of patients in a nursing home leads to important improvements in the patients' health and quality of life. Proper fall detection leads to improved response times, which in turn lead to dramatic improvements in the prognosis of recovery. Wandering itself can lead to distress, and is an important symptom of patients with dementia. A pragmatic approach is therefore required: for wandering detection fewer sensors can be used, meaning that a larger area can be monitored. For fall detection, which requires more precise pose estimation and where high recognition rates and few false positives are critical, a correct combination of sensors should be applied.

In the Balance-IT project we noticed that camera surveillance is always associated with privacy issues. In doing the experiments and gathering the data we were well aware of this. For the early work (presented in this paper) we used students to play the falls. In later work on fall detection that was carried out in a center for epileptic patients we used an actor (one of the care givers) who was extremely good in mimicking the epileptic attacks. .For the wandering detection we had a consent of the board of clients (clientenraad) to make recordings of the people with dementia and of, the head of the nursing institute to record the employees.

Acknowledgement

This work has been funded by the Foundation Innovation Alliance (SIA - Stichting Innovatie Alliantie), in the framework of the Balance-IT project. We thank AMSTA for their cooperation.

References

[1] Gallagher, B., Corbett, E., Freeman, L., Riddoch-Kennedy, A., Miller, S., Smith, C., Radensky, L., Zarrow, A. (2001) A Fall Prevention Program for the Home Environment, *Home Care Provider, **6**(5)*, 157–163.

[2] Gillespie, L. (2004). Preventing falls in elderly people: we need to target interventions at people most likely to benefit from them. *BMJ: British Medical Journal, **328**(7441)*, 653-654.

[3] Igual, R., Medrano, C., Plaza, I. (2013). Challenges, issues and trends in fall detection systems. *Biomedical Engineering Online, **12**(1)*, 66.

[4] Josemans, W., Englebienne, G., Kröse, B. J. (2013). Fusion of Color and Depth Camera Data for Robust Fall Detection. *Proceedings of VISAPP (1)*, p. 608-613.

[5] Liu, C.L., Lee, C.H., Lin, P.M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert Systems with Applications, **37**(10)*, 7174-7181.

[6] Martino-Saltzman, D., Blasch, B.B., Morris, R.D., McNeal, L.W. (1991). Travel behavior of nursing home residents perceived as wanderers and nonwanderers. *The Gerontologist*, ***31**(5)*, 666-672.

[7] Mastorakis, G., Makris, D. (2014). Fall detection system using Kinect's infrared sensor. *Journal of Real-Time Image Processing, **9**(4)*, 635-646.

[8] Rougier, C., Auvinet, E., Rousseau, J., Mignotte, M., Meunier, J. (2011). Fall detection from depth map video sequences. In *Toward Useful Services for Elderly and People with Disabilities* (pp. 121-128). Berlin/Heidelberg, Germany: Springer.

[9] Tao, J., Turjo, M., Wong, M.F., Wang, M., Tan, Y.P. (2005). Fall incidents detection for intelligent video surveillance. *Information, Communications and Signal Processing, 2005 Fifth International Conference on*, p. 1590-1594.

[10] Yu, X., Wang, X., Kittipanya-Ngam, P., Eng, H.L., Cheong, L.F. (2009). Fall detection and alert for ageing-at-home of elderly. In *Ambient Assistive Health and Wellness Management in the Heart of the City*, p. 209-216. Berlin/Heidelberg, Germany: Springer.

[11] Vuong, N.K., Chan, S., Lau, C.T., Lau, K.M. (2011). Feasibility study of a real-time wandering detection algorithm for dementia patients. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare*, p. 11. ACM Press.

Sampling the movement phenospace: Towards a biophysics of behavior in the wiggling of *C. elegans*.

Greg Stephens

VU Amsterdam
gjstephens@gmail.com

We apply a low-dimensional yet complete representation of body shape (eigenworms, [1]) to construct a principled parameterization of the 2D movement behavior of the nematode *C. elegans*. Despite it's simplicity, we show that a linear dynamical system of the eigenworm projections captures long-range temporal correlations and reveals two periodic dynamics, the primary body wave and and an oscillation between the head and body curvature which underlies long arcs in the organism's trajectory. We parameterize the movement phenospace by reconstructing equations locally in time and show that variation within this space is restrained; with increasing window size, a single behavioral mode dominates the variance and represents the coupled control of speed and turning. The distribution of this primary mode is bimodal, suggesting a correspondence with empirical behavioral states of roaming and dwelling. Finally, we apply our behavioral parameterization to show that the worm's response to a strong impulsive heat shock includes a Hopf-like bifurcation corresponding to an early-time growth of the amplitude of the crawling wave.

References

[1] Stephens, G.J., Johnson-Kerner, B., Bialek, W., Ryu, W.S. (2008). Dimensionality and dynamics in the behavior of C. elegans. *PloS Computational Biology*, ***4****(4)*, e1000028.

Syntax in *C. elegans* locomotion

André E.X. Brown

MRC Clinical Sciences Centre, Imperial College London, UK.
andre.brown@csc.mrc.ac.uk

Visible phenotypes have played a critical role in understanding the molecular basis of behaviour in model organisms. In previous work, we have shown that automated tracking and quantitative measurement of posture and motion parameters can sensitively identify new phenotypes in mutants of the nematode worm *C. elegans* [1]. However, it is not clear whether these parameters are in any way optimal nor whether they are useful for discovering new behaviours that we did not think to measure. As an alternative, we have also used unsupervised methods to automatically detect repeated behaviours in an unbiased way. A dictionary of such 'behavioural motifs' can be used to construct a phenotypic fingerprint for mutant strains and seems to be a promising way of predicting possible genetic associations [2].

Both of these approaches still leave several fundamental questions about behaviour unanswered. For example, how complex is worm behaviour? How long do we need to watch a worm before we have a good sample of its behavioural repertoire in a given condition? To address these kinds of questions, we have developed a simple method to capture the sequence of postural states of freely behaving worms and used it to derive a compact discrete representation of behaviour. Repetitive behavioural motifs are then identified in an unbiased way using a hierarchical dictionary-based compression algorithm to identify the syntax of worm behaviour (Figure 1). As opposed to the dictionary derived in our previous method, the compression-derived dictionary contains all of the repeated sequences the worm performed and is, in this sense, complete.

The dictionary learned for spontaneous behaviour in wild-type worms can be used to compress data from mutants or worms in different environmental conditions to measure phenotypic similarity. Dictionary comparison can also be used to assess inter-individual variation and the compositionality of behaviour, that is, the extent to which behavioural adaptation involves the creation of novel repertoire elements or the reuse of existing elements in novel sequences. Finally, the hierarchical nature of behaviour is built from the bottom up and may give insight into the evolution of behaviour: low-level behaviours that are used in many contexts may be evolutionarily more constrained than those controlling behavioural choices at higher levels.

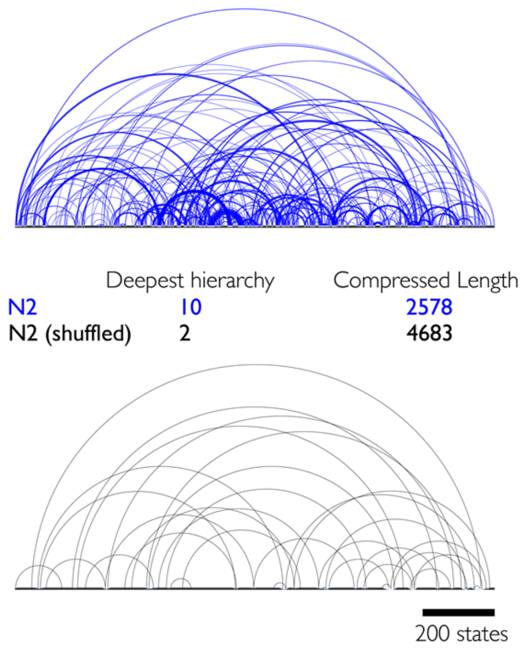


Figure 1. An arc diagram summarises the structure of worm behaviour by connecting segments of repeated behaviour identified in the compression-derived dictionary. Top: wild-type behaviour. Bottom: the same data after shuffling to destroy higher-order structure.

References

[1] Yemini, E., Jucikas, T., Grundy, L.J., Brown, A.E.X., Schafer, W.R. (2013). A database of Caenorhabditis elegans behavioral phenotypes. *Nature Methods,* ***10****(9)*, 877-879.

[2] Brown, A.E.X., Yemini, E., Grundy, L., Jucikas, T., Schafer, W.R. (2013) A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion. *Proceedings National Academy of Sciences USA,* ***110****(2)*, 791-796.

Locomotion in Tethered *Drosophila melanogaster* through Virtual Worlds

Tomás L. Cruz and Eugenia Chiappe

Sensorimotor Integration Lab, Champalimaud Neuroscience Programme

A fundamental goal of neuroscience is to understand how neural circuits transform sensory signals into sensory-guided behavior. Virtual Reality (VR) systems provide an experimental platform where animals perform naturalistic-like behaviors under precise control of sensory stimuli. In these systems, animals are usually restrained, and walk on the spot on moving treadmills, which, in turn, track the animal's locomotion and close the loop between behavior and the virtual environment to recreate an immersive experience [reviewed in 3]. Furthermore, restrained animal preparations are ideal for optically recording the activity of large populations of neurons in simultaneous with behavior [GCaMP6]. We recently developed *FlyVRena*, a VR system for tethered walking *Drosophila* that is based on a tracking system [4], a novel software platform, and a high-speed projection system [5]. Using FlyVRena, we studied visually guided locomotion in flies, an ethologically relevant task for this highly visual animal. Tethered flies walking on a spherical treadmill (an air-suspended ball) can perceived and interact with 3D-VR worlds as showed in spatial orientation tasks (Cruz *et al*, unpublished). However, because both the animal's restrained position and the moving treadmill itself constrain the locomotive behavior, it is difficult to make comparisons to natural walking conditions. Because our goal is to understand the link between neural activity and oriented walking, it is necessary to characterize those constrains, and analyze the consequent imposition they exert on locomotion. Here we present our attempts to characterize locomotion in tethered *Drosophila* walking on a ball through visual VR worlds. Using the dynamical signals from the treadmill tracking system, which measure the ball rotations, we first noticed different modes of ball motion induced by the fly. Notably, a subset of these modes correlates with high behavioral performance in orientation tasks, suggesting that they may represent naturalistic walking gates. To analyze the correspondence between ball rotations and walking under visual stimulation, we recorded the animal's locomotion using both the tracking system and videography [6]. Previous work has attempted to characterize leg movements with ball rotations [7]; however, this technique only detects the x-y position of the legs, which prevents recording other behaviors such as pushing and pulling of the ball, or abdomen movements. We therefore set two orthogonal cameras to measure leg and abdomen movement in 3D. Using this information, we developed a decoder to extract ball rotations associated with coordinated walking. Furthermore, we compared different kinetic parameters of free walking fly trajectories with those from reconstructed virtual trajectories in order to explore the effects of the constrains imposed by the VR system. The results we present in this work provide an unbiased description of the fly's maneuvering of the ball.

References

[1] Szuts, T.A., Fadeyev V., Kachiguine S., Sher A., Grivich M.V., Agrochão M., et al. (2011). A wireless multi-channel neural amplifier for freely moving animals. *Nature Neuroscience*, ***14****(2)*, 263-269.

[2] Thomas, S.J., Harrison, R.R., Leonardo, A., Reynolds, M.S. (2012). A battery-free multichannel digital neural/EMG telemetry system for flying insects. *IEEE Transactions on Biomedical Circuits and Systems*, ***6****(5)*, 424-436.

[3] Dombeck, D.A., Reiser, M.B. (2012). Real neuroscience in virtual worlds. *Current Opinion in Neurobiology*, ***22****(1)*, 3-10.

[4] Seelig, J., Chiappe, M.E., Lott G.K., Dutta A., Osborne J.E., Reiser M.B., Jayaraman V. (2010). Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. *Nature Methods*, ***7****(7)*, 535-540.

[5] Cruz, T.L. (2013). Development and test of a Virtual Reality system for tethered walking Drosophila. Master Thesis, Técnico Lisboa.

[6] Berman, G.J., Choi, D.M., Bialek, W., Shaevitz, J.W. (2013). Mapping the structure of drosophilid behavior. arXiv preprint arXiv:1310.4249v2.

[7] Kain, J., Stokes C., Gaudry Q., Song X., Foley J., Wilson R., de Bivort B. (2013). Leg-tracking and automated behavioural classification in Drosophila. *Nature Communications*, ***4****(5)*, Article No. 1910.

Deconstructing the value-based serial decision making process in *Drosophila* foraging

V.M. Corrales[1,2,3], A.A. Faisal[3,4,5], C. Ribeiro[2]

[1]Universidade Nova de Lisboa, MIT-Portugal Program, Lisbon, Portugal
[1]veronica.corrales@neuro.fchampalimaud.org
[2]Behavior and Metabolism Laboratory, Champalimaud Centre for the Unknown, Lisbon, Portugal
[2]carlos.ribeiro@neuro.fchampalimaud.org
[3]Department of Bioengineering, Imperial College London, London, United Kingdom
[4]Department of Computing, Imperial College London, London, United Kingdom
[4]a.faisal@imperial.ac.uk
[5]MRC Clinical Sciences Center, London, United Kingdom

We investigated how decision making is modulated by sensory information and internal state, by studying foraging behaviour in *Drosophila*. Flies balance yeast (main protein source) and carbohydrate intake depending on their metabolic and mating state to maximise fitness [1,2]. While proteins are required for offspring production they also reduce the lifespan of the animal. In order to understand the computations underlying this ethologically relevant value-based decision-making process, we decided to use a quantitative approach to capture the dynamics of this nutritional decision. We developed an automated machine-vision system, that records single fruit flies as they forage in a circular arena (radius = 3.3cm) containing 18 nutrient spots. Initially, 290 female flies were presented with two types of food options, 9 sucrose and 9 yeast spots. Each spot is composed of 18% food mixed with 0.75% agarose. As a control for non-food related activity, an additional group of 83 females was presented with 18 spots containing only 0.75% agarose. All flies' behaviour was recorded for 2 hours.

We first observed a characteristic speed profile that ranged between 0 to 4 mm/s, displayed only around food spots and not agarose spots. Based on this observation, we used both, distance to spot and speed, to determine the probability of engagement in a food spot at each point in time. The area below each engagement peak corresponds to the approximate duration of that engagement bout. We found that the distribution of sucrose durations follow power-laws characteristic of scale-invariant dynamics [3,4].

To assess the effect of internal state in the sequential dynamics of these decisions, we looked at the transition probabilities between spots and classified them into three types: transition to 1) a different spot of the same food type, 2) a spot of different food type or 3) the same spot (revisits). We found that mating has a very significant effect on both the sequential and temporal dynamics of these decisions resulting in a switch in nutrient preference from carbohydrates to proteins as captured by an increased total time in yeast ($p<0.001$) after mating. Currently, we are evaluating the effects of selective nutritional deprivations on these parameters.

Together these findings show that the internal state modulates the departure decision ("explore or exploit") and the arrival decision ("choose or move on") in order to achieve homeostasis. Our automated analysis makes these computational questions directly amenable to testing by genetic and neuronal manipulation.

References

[1] Ribeiro, C., Dickson, B.J. (2010). Sex Peptide Receptor and Neuronal TOR/S6K Signaling Modulate Nutrient Balancing in *Drosophila. Current Biology,* ***20****(11),* 1000-1005.

[2] Lee, K.P., Simpson, S.J., Clissold, F.J., Brooks, R., Ballard, J.W.O., Taylor, P.W., Soran, N., Raubenheimer, D. (2008). Lifespan and reproduction in Drosophila: New insights from nutritional geometry. *Proceedings of the National Academy of Sciences,* ***105****(7),* 2498–2503.

[3] Tavares, G., Faisal, A. (2013). Scaling-laws of human broadcast communication enable distinction between human, corporate and robot Twitter users. *PloS One,* ***8****(7),* e65774.

[4] Proekt, A., Banavar, J.R., Maritan, A., Pfaff, D.W. (2012). Scale invariance in the dynamics of spontaneous behavior. *Proceedings of the National Academy of Sciences,* ***109****(26),* 10564–10569.

Automatic high throughput measurement of feeding behavior in Drosophila

P. M. Itskov[1], J. Moreira[1], E. Vinnik[1], G. Lopes[1], S. Safarik[2], M. H. Dickinson[2] and C. Ribeiro[1]

[1]Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon, Portugal, pavel.itskov@neuro.fchampalimaud.org

[2] Department of Biology, University of Washington, Seattle, USA

Introduction

Feeding is an essential component of an animal's behavior and better understanding of the principles governing food intake will provide future insights into the interaction between physiology and behavior. Fruit flies (*Drosophila melanogaster)* have emerged as a powerful model to study the molecular and neuronal mechanisms underlying feeding behavior, but it remains challenging to quantify feeding in flies, due to their small size and the minute quantities of food they ingest. Existing methods are a potential limiting factor in the study of feeding in *Drosophila* because of their poor temporal resolution, reliance on manual annotation and inability to access the behavior of individual animals. Here, we have developed and validated an automated, high resolution behavioral monitoring system called flyPAD (fly Proboscis and Activity Detector), that uses capacitance–based measurements to detect the physical interaction of individual *Drosophila* with food.

Methods

The device uses ultralow current capacitance to digital converters to sample capacitance across 2 electrodes at 100 Hz and is capable of measuring feeding behavior of individually housed flies. To allow for high throughput behavioral analysis our system can record the behavior of up to 32 individual flies. We have furthermore developed an algorithm to automatically analyze the capacitance signal to extract behavioral features based on the shape of the capacitance trace. Our algorithm detects individual proboscis extension events ("sips") as well as periods of activity on the food ("activity bouts"). In order to validate our approach we have performed simultaneous capacitance and high resolution video recording.

Results

A comparison of the ethograms generated by manual annotation with the results of our automated method confirmed the accuracy of our approach. The number of "sips" detected by our algorithm was strongly and significantly correlated with the number of proboscis contacts with the food (Rho=0.874, $p<<0.0001$). The algorithm detected 92.5% of the sips tabulated via manual scoring, while missing 7.5% and generating 7.5% false sips. We have found that feeding from a non-liquid food induces a pattern of highly stereotyped rhythmic proboscis extensions and retractions that is suggestive of an underlying central pattern generator controlling the feeding motor program. By adapting the luciferase bioluminescent techniques to measure the intake in single flies we measured that 1 "sip" detected by the flyPAD corresponds on average to an intake of 1 nl of food. The analysis of ingestion dynamics and the microstructure of meals allowed us to dissect the behavioral elements mediating the homeostatic response of the fly to starvation and satiation. These results uncover several similarities with rodents and humans, highlighting a potential conservation of strategies that regulate food intake across phyla.

Conclusions

Our work opens up new avenues for studying the neuronal basis of feeding behavior and food intake at different time scales in an automated and high-throughput manner. The ability to subdivide homeostatic changes in feeding into specific changes in feeding strategies opens up new opportunities to explore how molecular mechanisms, neuronal circuits and metabolic processes interact to control food intake and nutrient homeostasis to optimize life history traits such as reproduction and aging.

Coordination of translation and rotation in insects: A behavioural homology with rodents?

Alex Gomez-Marin[1], Efrat Oron[2], Anna Gakamsky[2], Dan Valente[3], Yoav Benjamini[4], Ilan Golani[2]

[1] Champalimaud Neuroscience Programme, Lisbon, Portugal
[2] Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Israel
[3] Cold Spring Harbor laboratory, USA
[4] Department of Statistics and Sagol School of neuroscience, Tel Aviv University

Background

In previous studies it has been shown that in vertebrates the ground plan of locomotor behavior consists of a transition from extensive mobility to immobility through an increase and then a gradual decrease in forward translation, accompanied by a concurrent gradual increase in whole body rotation around the hindquarters (shut-down of behavior). The opposite transition from immobility to extensive mobility involves an increase and then decrease in rotation around the hindquarters, accompanied by a concurrent gradual increase in forward translation (warm-up) [1,2,3].

Aims

Here we examine the similarity between the mobility gradient exhibited in vertebrates and the shut down and buildup of mobility exhibited in fruit flies under the influence of cocaine. To accomplish this aim we characterize the dynamic relationship between rotation, turning, and progression, in fruit flies walking on a flat surface.

Methods

We use high-resolution computer-vision tracking of the animal's position and body orientation within a drug administration apparatus that controls the flow of cocaine via an evaporation chamber. The behavior, which is performed in a large circular arena (diameter of 150mm) is continuously recorded for an extended period of time (from minutes to several hours). Our analyses of the process include a novel phase-space representation of body rotation, translation and path curvature that illustrates the qualitative aspect of the phenomenon, presenting the moment-to-moment dynamics quantitatively.

Results

We discover an intricate interplay in the active management of these three degrees of freedom which, upon cocaine administration, exhibit a sequence that appears to correspond to the shut-down described in vertebrates. Specifically, progression decreases as rotation around the hindquarters increases and then subsides to complete immobility (Figure 1a). As the fly recovers, the inverse sequence of transition from immobility, to extensive rotation in place around the hindquarters followed by a gradual increase in forward translation (warm-up) is observed (Figure 1b).

Implications

Our results in flies thus bears an intriguing similarity to the mobility gradient exhibited in vertebrates suggesting that part of the mobility gradient is homologous in the two phyla.

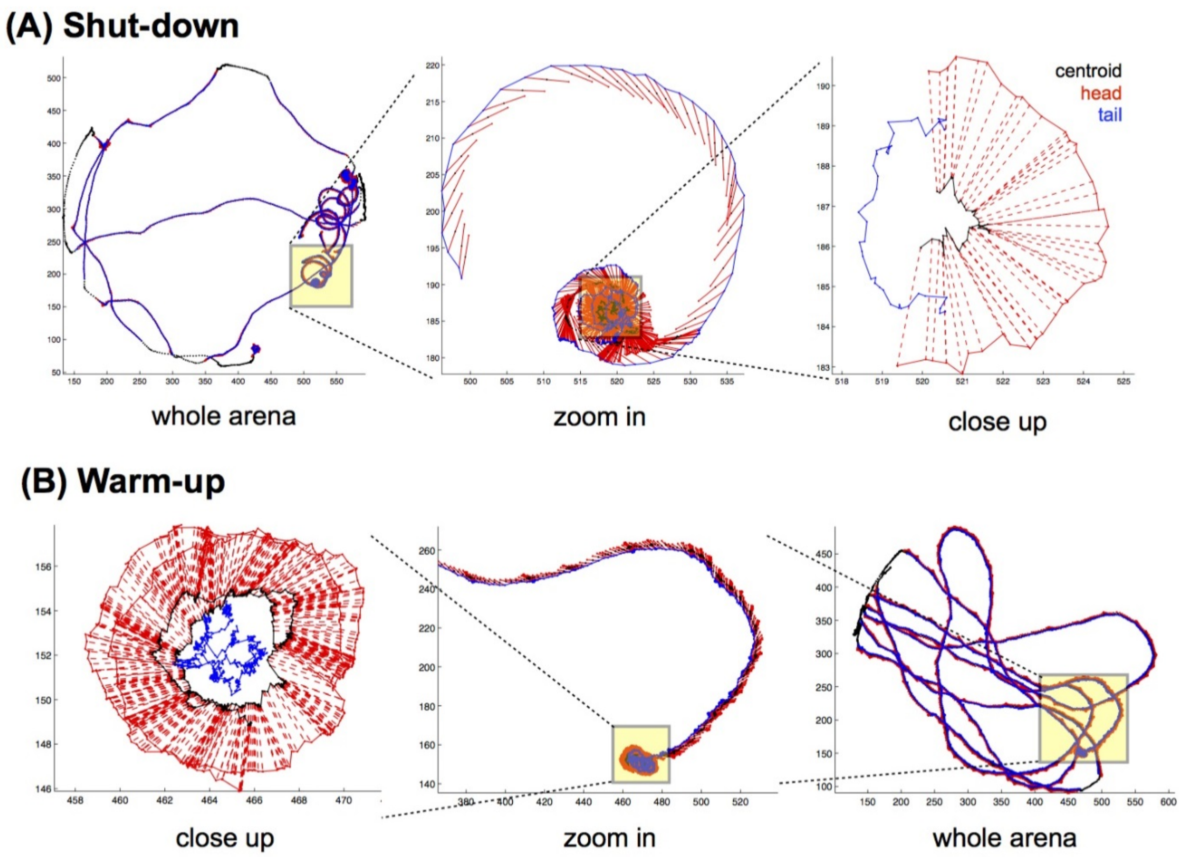


Figure 1. a. Shut-down: (Left) Trajectory of a cocaine treated fly in the arena. As the drug takes effect there is a transition from a path with low curvature to increasingly higher curvature until complete cessation of translation. (Middle) Zoom in reveals that in this case during this interval the front angle (heading direction) of the fly leads while the fly shifts its weight (direction of progression) maintaining a stable angular interval between path direction and front direction. In the last stage of shutdown, translation is eliminated and the fly rotates in place until full immobility. (Right) Further zoom in displaying a 180 degrees body rotation in place. As shown, the fly rotates around an axis located closer to the rear part of its body. The red, black and blue lines trace the head, centroid and tail trajectories. **b**. Warm-up: (Left) Following pronounced arrest as drug effect wears off, the fly proceeds with pure rotation around its hindquarters. This corresponds to the initial phase of transition out of immobility. The red, black and blue lines trace the head, centroid and tail trajectories. (Middle) Zoom out depicts the gradual increase in forward translation along a path whose curvature is gradually decreasing. (Right) As warm-up progresses, the fly resumes extensive forward locomotion increasingly resembling normal behavior.

Acknowledgements

This study was funded by the Israel Science Foundation (grant No 760/08).

References

[1] Golani, I. (1992). A mobility gradient in the organization of vertebrate movement: The perception of movement through symbolic language. *Behavioral and Brain Sciences*, ***15****(2)*, 249-266.

[2] Golani, I. (1992). The natural geometry of a behavioral homology. *Behavioral and Brain Sciences*, ***15****(2)*, 291-308.

[3] Golani, I. (2012). The developmental dynamics of behavioral growth processes in rodent egocentric and allocentric space. *Behavioural Brain Research*, ***231****(2)*, 309-316.

Quantifying exploratory behavior in mice: the vertical dimension

Yair Wexler[1], Ilan Golani[2], Liad Shekel[1] and Yoav Benjamini[1]

[1]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel
[1]yairwex@gmail.com
[2]Department of Zoology, Tel Aviv University, Tel Aviv, Israel

Introduction

Studies of exploratory behavior in rodents are conventionally based on tracking animal movement in a flat arena. Most studies of both forced and free exploration restrict the animal to horizontal locomotion. In this presentation we add the vertical dimension to the traditional 2D setup, and suggest how to quantify vertical movement. We focus on free exploration, where the animal is allowed to enter and leave the arena to an adjacent home cage, connected to it through a doorway, as it pleases.

A previous study on BALB/c mice behavior in a novel flat circular arena [1][2] provides evidence that this behavior consists of sequences of repeated motion away from and back to specific points of reference. In measuring this behavior in reference to a polar coordinate system whose origin is at the arena center, the angular and radial dimensions are explored via near-wall excursions from the doorway and incursions, which are performed from the wall towards center (Figure 1). The growth in both extent and complexity of these sequences has been extensively documented.

For example, the first excursions are straightforward and end with the animal leaving the arena after reaching a certain maximal angular distance from the entry point. However, as the maximal excursion distance from home is extended, more complex patterns appear. Excursions may now end in the vicinity of the doorway rather than in complete exit ("cage-skip"), or in other cases the mouse may reach a certain point and then backtrack part of the way before setting off again to extend the excursion further ("shuttle"). Shuttles such as this are also exhibited in incursions towards the center.

Quantifying movement in the vertical dimension

In the aforementioned and other open field setups, exploration of the vertical dimension was limited to rearing and jumping. However, surrounding the arena with a vertical wire grid on which the mouse can climb freely revealed sequences of repeated motion ("ascents") growing in a similar fashion to the horizontal ones (Figure 1). At first, the ascents are strictly vertical. The mouse climbs up a certain distance before descending all the way to the floor along the same route. Degrees of freedom are added to the motion on the grid as the maximal height per ascent gradually extends – by performing vertical shuttles, by progressing on the wire horizontally after reaching a certain height and later by climbing freely in diagonals involving vertical and angular change all around the grid.

Quantifying the growth in extent of the ascents (Figure 2) is based on methods previously used for quantifying near-wall excursions and incursions to the center [3]. The dependence of the maximal height per ascent on time (scaled by activity) is estimated using percentile LOESS smoothing. The comparative variables are the time to reach certain thresholds (e.g. 20% percent of the maximal height) and the growth rate at those thresholds, where growth rate is defined as the slope of the derivative at a particular point on the smoothed function.

So far this experiment was performed only with BALB/c mice, but considering the large differences observed in these two measures in pilot comparisons, we expect that significant differences will characterize ascents across strains, genders and preparations.

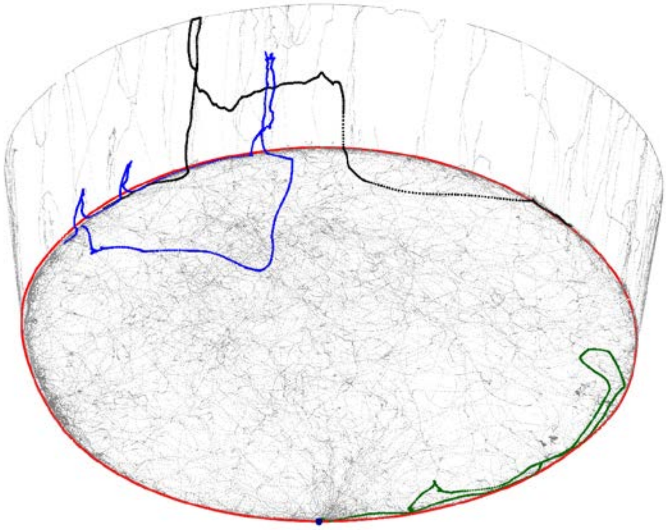


Figure 1. A 3D plot of exploratory patterns on the arena floor and the vertical grid. The arena boundary is highlighted in red, and the dark blue point marks the entrance. The three highlighted routes demonstrate addition of degrees of freedom to the mouse locomotion, sorted by time. Each is taken from a different entry. The earliest (dark green, starts from the left) is a simple near-wall excursion, starting and ending in the doorway. The second (blue, starts from the leftmost point and continues along the wall) includes three simple ascents growing in extent, followed by an incursion towards the center. The latest (black, starts on the right) demonstrates a more complex ascent, containing movement in the vertical and the angular dimensions, as well as a small vertical shuttle.

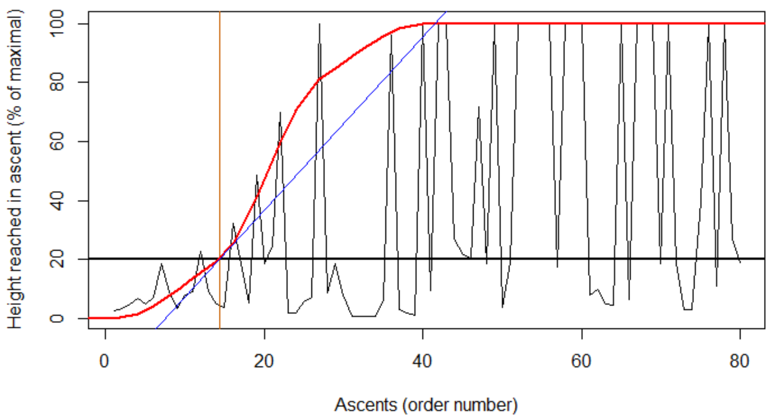


Figure 2. Quantifying the growth in extent in ascents. The red line represents the percentile-LOESS function. The black horizontal line is a certain selected threshold (in this case 20% of maximal height), and the vertical orange line is the time (in activity) to reach the threshold after smoothing. The blue line is the growth rate at the selected threshold.

Experimental setup, protocols and analysis

Except for the addition of the 47.5 cm tall vertical grid, the experimental setup and the protocols are described in details in [1]. 2D tracking was performed using Noldus EthoVision XT 9 [4]. The study was conducted in accordance with the Guide for Care and Use of Laboratory Animals provided by the NIH, ‘‘Principles of Laboratory Animal Care’’ (NIH publication no. 86–23, 1996).

Analyzing wall movement included the following phases:

- Converting the raw tracking data from a Cartesian to a polar coordinate system
- Defining accurate floor-wall boundary to isolate wall coordinates
- Smoothing wall coordinates separately
- Quantifying growth in extent of ascents

Considering the flat image produced by the single camera, analyzing the movement on the grid required an accurate separation of the grid from the floor. For that we use an algorithm which calculates the floor boundary based on estimates provided by the user. The user is required to draw a low resolution polygon on the mouse XY plot where the boundary should be, and the algorithm corrects deviations from the actual boundary and smoothes the polygon. The algorithm does not depend on arena symmetry and was found accurate for all kinds of arena shapes, from ellipses to rectangles, crosses and stars, as long as the boundary is clear to the eye (Figure 3).

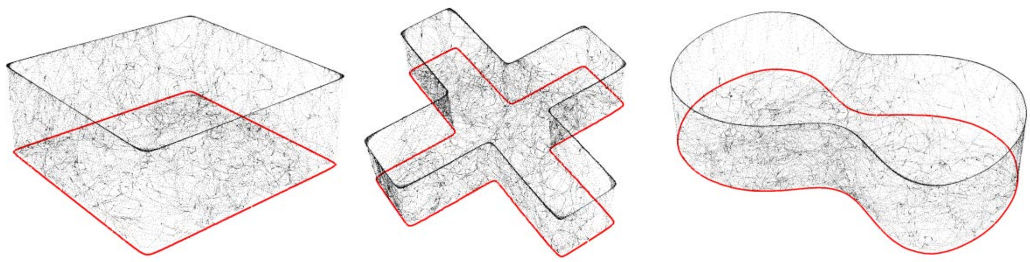


Figure 3. 3D images of mouse movement in different symmetrical and asymmetrical shapes of arenas, using one overhead camera and regular 2D video tracking.

Using the center of the arena as the origin of axes, the arena is divided to sectors by angle from the entry point, and the mouse XY coordinates are converted in the same manner to polar coordinates (r, θ). For each sector i, the separation function provides the boundary radius (R_i) per sector. Points within a specific sector found beyond the boundary $(r > R_i)$ are considered wall coordinates and are transformed to a cylindrical coordinate system (ρ, θ, z), where $z = r - R_i$ and ρ is held constant at R_i. The θ and z coordinates are then smoothed using LOESS algorithm.

Ascents are defined as periods of time between the first and last time stamp in which $z > 0$. They are quantified as explained before. Growth rate and time to reach threshold in two groups of mice may be compared using unpaired two sample t-test.

Floor coordinates $(r \leq R_i)$ are converted back to XY and the movement is analyzed, including position smoothing, retrieval of velocity and acceleration and motion segmentation, in the SEE software [5] (http://www.tau.ac.il/~ilan99/see/help/).

Summary

Conventional horizontal tests of locomotion and exploratory behavior in rodents are limited, and therefore may fail to detect important differences between groups. According to this hypothesis, we suggest addition of vertical locomotion to these tests, as well as methods to quantify it. We believe that this addition may reveal patterns unique to different strains and conditions. The techniques presented enable adding a vertical dimension to any former test setup, without the additional cost and restrictions of a 3D tracking system which requires more cameras and relatively small tanks.

References

[1] Fonio, E., Benjamini, Y., Golani, I. (2009). Freedom of movement and the stability of its unfolding in free exploration of mice. *Proceedings of the National Academy of Sciences,* ***106***, 21335-21340.

[2] Golani, I. (2012). The developmental dynamics of behavioral growth processes in rodent egocentric and allocentric space. *Behavioral Brain Resaerch,* ***231***, 309-316.

[3] Benjamini, Y., Fonio, E., Galili, T., Havkin, G.Z., Golani, I. (2011). Quantifying the buildup in extent and complexity of free exploration in mice. *Proceedings of the National Academy of Sciences,* ***108***, 15580-15587.

[4] Spink, A.J., Tegelenbosch, R.A.J., Buma, M.O.S., Noldus, L.P.J.J. (2000). The EthoVision video tracking system: a tool for behavioral phenotyping of transgenic mice. *Physiology & Behavior,* ***73***, 731-744.

[5] Drai, D., Golani, I. (2001). SEE: A tool for the visualization and analysis of rodent exploratory behavior. *Neuroscience & Biobehavioral Reviews,* ***25***, 409-426.

Quantifying continuous behavior to account for neural variability

A. Motiwala*, T. Gouvêa, T. Monteiro, J. Paton[1], C. Machens[1]

Champalimaud Neuroscience Program, Lisbon, Portugal
*asma.motiwala@neuro.fchampalimaud.org

A central question in investigating neural function is to understand the nature of the computations being implemented in different neural circuits. To address this question we need to understand what exactly populations of neurons encode. In experimental neuroscience, neural activity is usually recorded in highly controlled contexts where different aspects of the environment are systematically varied. Neural responses are then studied in response to these systematic experimental manipulation. However, some aspect of neural responses are likely to vary with additional uncontrolled parameters such as the animal's continuous behavior or movement states. Here, our goal is to develop an approach to quantify the dynamics of an animals' behavior. This will allow us to investigate if there is any structure in neural dynamics that could be explained by the structure of behavior during the task.

There are two key aspects to consider in quantifying behavior for this purpose— first, what is an appropriate, unbiased measure of behavior? And second, what is the most succinct yet minimally lossy representation of the measured behavior? The problem of representation is usually not an issue when only a few variables are recorded since the original measures can be directly used for analysis. But, measuring only a few variables may fail to capture important features of animal behavior. On the other hand, when the number of variables is increased, it becomes harder to characterize the trends in the data. Hence, an appropriate low-dimensional representation of the high-dimensional data is required.

With regard to the first issue—conventional techniques of obtaining continuous measures of behavior involve identifying and tracking some set of features of an animal's body, such as the position of the animal's centre of mass, head position, paw position, orientation, etc. These features can be combined to define different behavioral states of the animal. This allows a clear interpretation of relationships between well defined behavioral states and relevant environmental or task related variables. However, it suffers from several image processing problems in extracting relevant features and, more profoundly, from the bias introduced by the experimenter in the choice of features that are tracked. To overcome these, we directly use high speed videos of animals performing a stimulus categorization task as an unbiased measure of the animal's ongoing behavior.

This brings us to the second issue. In working directly with high speed videos, we are treating each pixel in the video frames as a continuous time-varying variable. Clearly, this is an extremely high-dimensional representation. Especially when high resolution images are collected. Hence, the task of determining an appropriate low-dimensional representation is paramount. Since we are using a top view of an animal performing a task in a box, we may be justified in assuming that the intrinsic geometry of the data is likely to be rather low-dimensional. The different dimensions of this underlying structure would correspond to translation, rotation, and changes in the contours of the animal's body as it performs the experimental task. There are several analysis methods with which high-dimensional data can be compactly represented. Among the most commonly used are linear unsupervised techniques such as principle component analysis (PCA). However, the structure in images of animals is likely to be highly non-linear in pixel space and linear techniques, such as PCA, may be inadequate in pulling out any underlying low-dimensional structure. Hence, we are investigating the efficacy of a non-linear dimensionality reduction technique called Isomap [1] to represent videos of behaving animals.

Dimensionality reduction procedures such as PCA can be understood in terms of two steps—1. Determine a measure of similarity between all data point. In the case of PCA, this would be covariance. 2. Find an arrangement of data points in a low dimensional space such that it preserves the measured similarity between the data points. Isomap is implemented using a similar procedure. The key distinguishing point, however, is the choice of the similarity metric. Isomap uses an approximation of Geodesic distances between data points. The second embedding step finds an arrangement of points in a low dimensional space in which the Euclidian distance between the points approximates the Geodesic distances computed in the original high dimensional space. This procedure unfolds curved manifolds to reveal their true underlying dimensionality.

[1] Equal contribution.

Isomap has been previously used for analysis of image sequence [2]. The presented work is an attempt to investigate its relevance as a technique to represent videos of behaving animals. We describe the nature of high-dimensional representations, in pixel space, generated by different transformations of objects in images. Subsequently we show how well the representations extracted by linear and non-linear dimensionality reduction techniques capture the structure of the original data. Some of the considerations in implementing Isomap, that affect the result of the analysis, are discussed, as well as some of its limitations. Furthermore, using toy data as control reference, we discuss the appropriateness of this method to work with videos of behaving animals and what criteria do the data need to satisfy for Isomap to be a suitable technique to obtain a low dimensional representation.

Ethical statement

All experiments were approved by the Champalimaud Foundation Bioethics Committee and the Portuguese Direção Geral Veterinária. The animals were allowed ad libitum food and access to bedding and sheltering material. They were maintained in a 12 hr-12 hr light-dark cycle. And to create motivational drive for the experimental task, they were water deprived to a maximum of 80% ad libitum body weight.

The data discussed in the presented work involves rats that were trained on a classification task. They were required to classify the duration of the delay between two tones as being either longer or shorter than a reference duration determined by the experimenter. This duration was not explicitly available to the animal but was learned by reward contingency based on choices in response to the different durations presented. Detailed descriptions of the experimental setup and paradigm are available in [3].

References

[1] Tenenbaum, J.B., De Silva, V., Langford, J.C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, **290**(5500)*, 2319-2323.

[2] Pless, R. (2003). Image spaces and video trajectories: Using isomap to explore video sequences. In *Proceedings Ninth IEEE International Conference on Computer Vision, Vol. 2*, p. 1433-1440. October 13-16, Nice, France. Los Alamitos, CA, USA: IEEE Computer Society.

[3] Gouvêa, T.S., Monteiro, T., Soares, S., Atallah, B.V., Paton, J.J. (2014). Ongoing behavior predicts perceptual report of interval duration. *Frontiers in neurorobotics, **8***, 10.

From 2D to 3D in ATC? Can operators be trained?

C.H.M. Nieuwenhuis [1], P. Chopin [2]

[1] Technical Directorate of Thales Nederland B.V.
[2] Direction Technique of Thales-Raytheon Systems, France

Abstract

Currently, ATC operators work with 2D representations of the airspace and air-traffic that they monitor and control. This controlled airspace is visualized on computer screens as a 2D map and textual information in the form of labels or pull down menus is available as supporting information. From this the operators have to build and constantly update a 3D model of the airspace in their head, which they use to validate their control decisions. Most of these decisions have to do with resolving location conflicts in time and space. With the predicted increase in air traffic in 2025 not only the workload of the operators will increase, but simple measures such as more operators simultaneously for smaller segments of airspace will not help. Therefore, new methods and techniques are needed to reduce the operator workload and automate more of the traffic control, while maintaining the same strict safety margins. One direction in which a solution could lie, is to use 3D Virtual Environments as the working space for the ATC operator, assuming that what is essentially a 4-dimensional control problem can be more easily and directly visualized and recognized in 3D (or time-enhanced 3D), thus freeing up cognitive processing capacity for the actual solution finding.

Historical perspective and still unanswered questions

The use of 3D techniques have been investigated before in the ATC domain, but not to the extent of actually using a 3D immersive virtual environment, specially constructed for doing controlled usability experiments and measuring operator performance and to investigate and experiment with the human factors and visualization aspects of a 3D VE based solution. The real question is if the use of a 3D virtual environment will actually help ATC operators to perform better with a more realistic and immersive rendering of their work domain, and in particular if they can be trained to develop a problem solving behavior that matches well with their perception capabilities in such a 3D visualization. And of course we have to find out if it is at all possible and acceptable for experienced operators to switch from their current solutions to such a new and break-through and significantly different solution.

To illustrate the sort of possible scenarios that drive new solutions for air traffic control, consider the following: the current practice in air traffic control is to line up the planes that come in to land in a straight line and down to the landing lane. Doing it that way requires a lot less cognitive resources from an operator to monitor for conflicting situations and solving them, then if planes were allowed to come from all directions and make their final approach with a curved path, coming from multiple directions 'at almost the same time'. But in order to cope with the expected increased traffic density and improve the air side capacity of an airport, the need for more air space will maybe drive a change in approach procedures that actually may give pilots the freedom to fly in from multiple directions for a curved final approach. In this case the work of a human air traffic controller to maintain correct situation awareness and spot and resolve conflicts in time, will become a lot more demanding.

Tools for usability and user performance studies

The 3D virtual environment should be able to support the two main ATC processes: en-route monitoring and control and take-off and landing monitoring and control, using 'life' data from sensor and other sources. The first process is probably less impacted, even under the higher workload conditions expected in future denser air traffic. But for the latter processes 3D is expected to make a significant difference in helping to handle the expected changes in traffic load and procedures. Probably the hardest task in designing the right 3D virtual environment set-up will be to determine what and how we need to measure and how to generate and capture all the different data streams that we need to set-up and analyze the experiments, with the many human operators in a repeatable and reliable way. From a cognitive and psycho-physiological perspective we want to measure things such as information overload, problem solving strategies, change blindness, vigilance, fatigue, and many more parameters.

And we want to measure them in relation with the many different events in the many complex scenarios. And we want to do experiments in VR looking at the performance of trained operators under controlled conditions. For this we need to build a dedicated experimental immersive VR environment, focusing on an ATC context, both civil and military. We will use tools and technology from third parties of course, but we will have to set up and validate our own experimental methods and capture and analysis techniques and tools. And not only do we want to measure the behavior of people during training exercises, to get a better understanding on how well they learn and what we can do to improve their learning curve, but we also need to find out how 'stable' their new skills are and how well they will transfer into the real work domain. The latter may sound trivial if the aim is to use a 3D VE technology as the new ATC system solution, but we cannot simply take that for granted.

And because the number of aspects that need to be looked at and understood is too big to handle in any reasonable amount of time by one institute or company, we need to set up partnerships with universities and SMEs to help in the work. Therefore, the 3D VE and the utilized measurement and analysis techniques, but also the scenarios and experimental set-ups need to be standardized, so that they can be build and used in several places and their results can be combined and integrated and become a corpus of re-usable experimental data.

Some early experiences with setting up and using such an immersive environment

With the help of a 3D cave-like technology provider (TechViz, in France), we have constructed a first prototype ATC environment and presented this on the Thales innovation Days and on Eurosatory 2014. These presentation served to familiarize Thales higher management and customers alike about the intended studies, to 'test the water' so to speak. These interactions also helped to generate many concrete questions about for example interfacing and visualization issues. And they generated new ideas about remote collaboration between controllers that could add a whole new dimension to 3D VE based solutions. A selected sample of these preliminary findings will be discussed in this presentation. The next step is to design and construct a 3D Virtual ATC Experimental Platform with a minimum of two and possibly more nodes, to facilitate distributed and remote collaboration between experimental traffic control centers. The Platform will be equipped with cameras, microphones, polygraphs, eye-trackers and other non-invasive and invasive measurement equipment to capture and register multiple simultaneous data streams, necessary to analyze various human factors aspects of air traffic controllers in experiments. Aspects such as workload and cognitive workload, cognitive control, stress factors, situational awareness, vigilance will be monitored with the help of existing and new-to-develop models that link the data to the human factor aspects of interest. Experiments will be designed (and must of course be validated) to test both operator and operational system performance. Special care must be taken to design test scenarios that allow the comparison of performance measures in both the current 2D situation and new 3D proposed ATC solutions. Baseline for the design of these experiments will be a thorough analysis of the current state-of-the-art in 2,5D, 3D visualization studies. See as an example of this literature [1], [2], [3]. Also recent work on navigation and selection of data in 3D scientific and medical data spaces will serve as a starting point for these future ATC directed studies. See as an example of this literature: [4], [5].

And finally, two aspects must be considered above all. The first is that in a safety critical domain such as ATC a new solution that is totally dependent on advanced computer technology, must always have some back-up solution that can be called upon. And the second is that a new solutions is only acceptable if it (provable!) makes air traffic safer than it already is today. This safety and performance behavior of the human-machine solution is of primary importance, but it is still an open question how to measure it and maybe transform it into a norm. What is clear is that the later can only be defined properly in an open discussion between end –users, scientists and industry.

References

[1] Tavanti, M. (2004). *On the relative utility of 3D interfaces*. PhD dissertation, Faculty of Social Sciences, Uppsala University.

[2] Cooper, M.D., Fridlund, A., Andel, M., Bojan, C., Hardy, J.-L. (2010). Educational benefits of 3D displays in early controller training. In I. Grant (Ed.), *Proceedings of the 27th International Congress of the Aeronautical Sciences (ICAS2010)*, Article No. 3. September 19-24, Nice, France. UK: Optimage Ltd.

[3] Bourgois, M., Cooper, M., Duong, V., Hjalmarsson, J., Lange, M., Ynnerman, A. (2005). In S. Saunders-Hodge and C. Pusch (Eds.), *Proceedings of the USA/Europe Seminar on Air Traffic Management Research and Development*, Article No. 40. June, 27-30, Baltimore, Maryland, USA.

[4] Isenberg, P., Isenberg, T., (2013). Visualization on interactive surfaces: A research overview. *i-com, **12**(3)*, 10-17.

[5] Klein, T., Guéniat, F., Pastur, L., Vernier, F., Isenberg, T. (2012). A design study on direct touch interaction for exploratory 3D scientific visualization. *Computer Graphics Forum, **31**(3)*, 1225-1234.

Studying the features of collaboration in the VirCa immersive 3D environment

B. P. Hámornik[1], A. Komlódi[2], M. Köles[3], K. Hercegfi[4]

[1,3,4] Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Budapest, Hungary
[1]hamornik@erg.bme.hu, [3]kolesm@erg.bme.hu, [4]hercegfi@erg.bme.hu
[2] Department of Information Systems, University of Maryland, Baltimore County, Baltimore, United States
[2]komlodi@umbc.edu

Abstract

Collaborative examination and use of shared documents often happens in a common physical space like in a tourist office. We developed a 3D immersive space to study collaboration on an information task in a virtual environment. In our paper, we present the qualitative results about cooperative work in the VirCA virtual environment. Our research is aiming to show that a 3D virtual environment can appropriately support collaborative information interpretation and sharing activities.

Keywords: Collaboration, Information search, Virtual Reality Cave, Observation

Introduction

Information-rich collaborative tasks can greatly benefit from tangible representations of information that the collaborators can easily share. However, sharing a physical space is often difficult and expensive when collaborators are distantly located. We set out to address this problem by developing a collaborative immersive 3D space with shared digital representations of documents.

We chose immersive 3D environments over 2D solutions for several reasons. First and foremost, we feel that 3D immersive environments can best replicate and improve on the advantages of shared physical spaces. Second, immersive 3D environments allow large amount of information to be displayed. 2D displays can also create high-density information displays, however, we assumed that more realistic 3D immersive environments would make it easier for users to browse, manipulate, interpret and use the information [5]. Third, a shared virtual space where both users are represented through avatars can help the users create common ground and would bring emotional commonality.

Shared virtual reality systems offer remote collaboration to geographically dispersed collaborators. Not only hearing but also seeing another person (or at least an avatar representing that person) will create a higher sense of presence, interpersonal trust and perceived communication quality [2]. The more realistic is the environment is the more effective is the spatial search [5,6]. Such collaborative virtual reality environments have been developed for various purposes. Among others, immersive 3D virtual environments can afford the shared viewing and manipulation of digital representations of documents in a common virtual space. An example of these systems and closest to our focus is VR VIBE [1], a collaborative virtual reality system in which users can browse and search web content in a 3D immersive space while at the same time seeing other users in the space using the same information. While we did not locate studies on collaborative information seeking tasks in immersive virtual reality, Raja and colleagues [7] found that immersive environments support individual information visualization tasks well. Our virtual space is similar to VR VIBE in its purpose and setup, however, our space supports the second half of the information process: the interpretation and use of information. Three important elements of our space are: 1) physical representation of information in posters; 2) user actions to manage these posters; 3) collaborative editing surface to create a new document based on the interpretation of the information contained in the posters. Typical actions necessary for the interpretation of are: reading, structuring and organizing, highlighting, commenting, and creating new content.

While immersive 3D environments have their challenges for users, such as disorientation and fatigue [8], they afford interaction patterns that are similar to shared physical spaces and thus they have the potential to support collaborative information tasks well. User interaction methods and information displays possess more flexible characteristics than physical spaces. In addition to physically moving and organizing information items, users can take advantage of digital capabilities, such as full text search, digital annotations, and different access levels for different users, exporting and importing digital formats, and so on. Another advantage of these spaces is that the number of documents, or information items, is theoretically limitless. The flexibility of the virtual environment allows users to browse, view, move, group, and annotate a large number of documents. While this is a great opportunity for users, designers have to carefully create the environments to make interaction natural and address the challenges of disorientation and fatigue for users of 3D immersive spaces.

The task and the environment are analogous to the real-life situation where travellers have to plan a trip in a tourist office. In a tourist office, information is displayed on the walls around the room in the form of brochures and posters. Tourists have to select, collect, and organize these pieces of information in order to make decisions based on them. Similar to a real tourist office, the virtual environment in VirCa (Virtual Collaboration Arena) [10] consists of the following: (1) Posters containing information on restaurants, sights, and events (see Figure 2 and 3). This information includes opening hours and addresses that are essential for the planning task. (2) A city map with numbers denoting the location of the restaurants, sights, and events from the posters. (see Figure 2). (3) Sticky notes that are note windows available for both collaborators to write in and attach to posters. (see Figure 3). (4) A jointly editable document where participants described their planned tour schedules. (see Figure 4). These objects were placed in a 3D room model that was accessed by the two collaborating participants. One participant was placed in a 3D Cave thus experiencing an immersive environment. The other participant joined the partner though a desktop computer that provided a less immersive environment. This mixed structure of Virtual Reality interfaces were designed to compare and analgise the features of information usage of collaborators and usability of interfaces in the future research similar to Tan, Gergle, Scupelli, & Pausch [9].

Figure 1. Collaborative review of the posters in the VirCA 3D cave.

Figure 2. Avatars in the VirCA 3D cave.

Figure 3. A sticky note in the VirCA 3D cave.

Figure 4. Collaborative editing in the VirCA 3D cave.

In this study, we aim to examine whether and how the 3D immersive environment can support a collaborative information management and use task analogous to a real world problem of planning tours of a city [3] in tourist office. This task serves as an example of joint document viewing and interpretation activities. In this paper we report the first descriptive results from an on-going larger study. As part of this research we are also exploring the usability of collaborative 3D virtual spaces based on our earlier work [4].

Method

Forty students (20 pairs) participated in the study. Participants were paired randomly; there was no matching of gender or age. In each pair, one participant was seated in front of a desktop personal computer (equipped with a Tobii T120 eye-tracker for further analyses not reported here), while the other was in the immersive 3D virtual cave.

Video recordings

Three video recordings were created from each session. One recording was a screen capture of the desktop participant's view. Another recoding was from an external video camera capturing the other participant's interactions in the immersive space. A third virtual and invisible "cameraman" was also used. This view was an invisible participant set up at a second desktop personal computer with a view into the virtual space. The viewpoint and thus the recording camera angle were controlled by one of the experimenters and were dynamically moved in the room to provide the best view on the interaction between the two participants. All three recordings contain the audio of the participants' conversation. The third recording was selected for analysis purposes as it captured the movements of both participants in the space.

Task and procedure

After participants arrived at our location, they were introduced to the goals and procedures of the study and they signed consent documents. Next, they received a brief introduction to the virtual reality space and the interaction tools at their disposal. They had a chance to ask any questions they wished and practiced using the space for a few minutes. Before starting the task, the connection between the two locations (one with the desktop computer and the other with the immersive cave) was established and the participants were introduced to one another. The task in the study was to plan a schedule of tours for a foreign student group spending a weekend in Budapest, Hungary. This task is similar to the problem discussed by Crabtree and colleagues [3]. In order to create the plan, participants had to use information (tourist attractions, restaurants, opening hours and location) displayed on posters that were placed on the walls of the virtual environment (Figure 1, 2, 3, 4) The detailed instruction of the constraints of the two-day trip schedule was also placed on the wall of the virtual room as a reminder. The plan itself had to be written in a shared editable document in the immersive space (Figure 4). This task is natural and familiar for students, as they confirmed in their post-interaction interviews. The participant in the 3D cave had a Lenovo N5902 mouse and keyboard combined tool as an input device. The task started with an instructed practice phase in order to become familiar with the environment and the functions. Than after the task completion came that had no time limit. Right after participants declared that they have finished the task, they were interviewed about their experience. Participants also completed a test of mental rotation (Paper Folding Test), filled out a demographic and gaming/virtual reality experience questionnaire, and two satisfaction questionnaires, one about the quality of the collaboration and another about the usability of the space. Finally, participants were asked to report their familiarity with Budapest and especially the locations presented in the posters. As described above, in each pair, one participant completed the task in an immersive environment, while their partner at a desktop computer. The two participants could see each other as an avatar in the space (Figure 2) and communicate verbally through headphones. The avatar was represented as a head and an adjustable arm (Figure 4). The ethical committee of the university approved the experiments.

Qualitative results and discussion

Our initial review of the observation notes and the video recordings resulted in a qualitative description of the collaborative behaviours exhibited by our participants and a high-level usability analysis. These results are reported here as foundations of further structured analyses.

There were cooperative activities around three elements of the 3D cave: (1) around the map for determining the location of the sights and events proposed; (2) around the posters while collecting information, including collecting possibly interesting posters as it can be seen on Figure 1; (3) and around the editable document where the participants shared the task of writing the final schedule (Figure 4). These three elements were crucial to the problem at hand and our participants decided to collaboratively interacting with the elements by viewing them at the same time and using their avatar arms to point at the objects. These patterns demonstrate that even the limited avatars allowed for a shared experience similar to a real world physical space. These results are the first indications that this 3D virtual space indeed allowed effective collaborative management and use of documents.

Interestingly the participants did not use the highlighting function in order to emphasise certain posters, e.g., by priority. Instead they used spatial location to organise the events and sights: they usually collected the selected posters in one place and discussed them afterwards. The highlight feature is redundant because the avatars are able to communicate synchronously in real time by auditory and visual channels. Pointing and mentioning an object is much more natural than highlighting it.

One of the most frequent usability issue participants had was tied to typing. Most participants reported that typing seemed difficult and immersion breaking. Participants both in the cave environment and at the desktop computer reported this. Thus, this finding was not uniquely related to the wireless keyboard device, although that device was not familiar for most of our users. The issues regarding typing point to differences between virtual and real world interactions. In a virtual environment, the need to look down at a keyboard is extraneous and immersion breaking. One solution could be providing a virtual keyboard, where participants would be able to type with their hands. The other direction would be based on the usual type of interaction participants displayed in the virtual space, and that is direct graphical manipulation. The posters already contain all the information they possibly know about a certain event. Writing this information out into another medium is most likely not the most efficient way. Instead having them organize the posters themselves into a two-day plan might be less immersion breaking and more natural in a visual environment. In general we can conclude that the environment reached that level of resemblance to reality that could serve the collaborative problem solving. According to the findings of Meijer and colleagues [5] the highly realistic virtual environment supports navigation and tasks inside them we conclude that the tourist office setting was realistically modelled in the VirCa. This is supported by the interviews also where none of the participants mentioned that the environment was hindering, or unrealistic.

Conclusions and future work

In our paper, we presented preliminary qualitative results about cooperative work in the VirCA virtual environment that mixed the cave and the desktop interfaces in this phase. We found that users successfully completed a collaborative information management and use task and exhibited effective collaborative behaviours. We identified three usability problems and described possible solutions to them. In the further phases of experiments in VirCa environment we may apply these solutions. Through these results we have shown that collaborative information seeking and used tasks can be successfully supported in shared virtual environments. We plan to provide further and more detailed evidence in our continued analysis of the data based on eye tracking analysis and sequential analysis of actions observed.

Acknowledgment

The authors would like to acknowledge the support provided by the European Union Research Executive Agency Marie Curie International Incoming Fellowship program via Grant Number 301974: Human-Information Interaction in 3D Virtual Environments (HII3D).

References

[1] Benford, S., Greenhalgh, C., Rodden, T., Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM*, ***44****(7)*, 79–85.

[2] Bente, G., Rüggenberg, S., Krämer, N.C., Eschenburg, F. (2008). Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations. *Human Communication Research*, ***34****(2)*, 287–318.

[3] Crabtree, A., Tolmie, P., Rouncefield, M. (2013). How many bloody examples do you want? Fieldwork and generalisation. In *Proceedings of the 13th European Conference on Computer Supported Cooperative Work*, p. 1–20. Springer.

[4] Hercegfi, K., Köles, M., Tóvölgyi, S., Komlódi, A., Józsa, E., Rózsa, G., Boross, J. (2011). Eye-tracking based usability evaluation of an emotional display object integrated to the virtual collaboration arena. In *Proceedings of the CogInfoCom 2011*, p. 5.

[5] Meijer, F., Geudeke, B.L., Van den Broek, E.L. (2009). Navigating through Virtual Environments: Visual realism improves spatial cognition. *Cyberpsychology & Behavior*, ***12(5)***, 517–521.

[6] Pausch, R., Proffitt, D., Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, p. 13–18.

[7] Raja, D., Bowman, D., Lucas, J., North, C. (2004). Exploring the benefits of immersion in abstract information visualization. *Proc. Of IPT (Immersive Projection Technology)*.

[8] Sebrechts, M. (2005). Visualizing information in virtual space: Prospects and pitfalls. In *Knowledge and Information Visualization*, p. 136–166.

[9] Tan, D.S., Gergle, D., Scupelli, P., Pausch, R. (2006) Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction*, ***13****(1)*, 71–99.

[10] Vámos, Á., Fülöp, I., Reskó, B., Baranyi, P. (2010) Collaboration in virtual reality of intelligent agents. *Acta Electrotechnica et Informatica*, ***10****(2)*, 21–27.

Design and evaluation of HMI concepts for cooperative driving through a driving simulator

Q. Shahab, J. Terken, B. Eggen

Department for Industrial Design, Technische Universiteit Eindhoven, Eindhoven, the Netherlands
{q.m.shahab,j.m.b.terken,j.h.eggen}@tue.nl

Abstract

Most developments for cooperative driving focus on automated systems taking over control from the driver (also known as Cooperative Adaptive Cruise Control), but it may take quite some time for such technology to achieve a sufficient degree of penetration in the market. Therefore, it has been argued that it might be worthwhile to develop aftermarket advisory systems for cooperative driving for the transition phase. In the current paper we summarize several studies focusing on the design and evaluation of an advisory system for cooperative driving. We show how such systems can be developed through studies employing a driving simulator and discuss where we hit the boundaries of what can be studied through a driving simulator.

Introduction

Cooperative Driving refers to vehicles driving in road trains (platoons) with short inter-vehicle time-gaps, where the acceleration and braking behaviour of a lead vehicle is copied to the following vehicles by means of wireless communication. It may be combined with automatic lane-keeping to provide a high level of automation for the following vehicles. Cooperative Driving has been claimed to provide benefits for highway traffic: it is expected to result in a more stable traffic flow, reducing the number of traffic jams following from unstable traffic flow ("phantom" traffic jams). Also, because of the small inter-vehicular distances, the road capacity is expected to increase. Finally, because of a more stable traffic flow and the reduced inter-vehicular distances, fuel consumption is expected to decrease. While first road tests with cooperative driving have shown the technical feasibility of cooperative driving, the technology is not yet mature and it is expected that it may take quite some time before the technology can be rolled out. Also, legislature needs to be modified, which again may take quite some time. Therefore, it has been suggested that advisory systems might provide a means to speed up the market penetration for cooperative driving. In assistance systems, the driver receives advice about the desired speed but needs to perform the required acceleration and braking actions himself. While this means that the inter-vehicular distances cannot be as small as with automated acceleration and braking, still it has been shown that such advisory systems already have a beneficial effect on traffic flow 1,2.

The aim of the research presented in this paper was two-fold: to investigate how to present speed advice to drivers (Interaction design) and to investigate how drivers may be persuaded to use such technology by applying persuasive technology. In the remainder of the paper we summarize three studies, which were conducted with a driving simulator in which platoon driving was implemented. The driving simulator setup is shown in Figure 1. Since cooperative driving is primarily intended for a highway context, all studies involved highway scenarios. At the end of the paper we will discuss methodological issues associated with conducting research on the topics under discussion with a driving simulator.

Ethical issues: Since only behavioural measures were collected and no recordings were made involving personal data, a low ethical risk level applied and no approval of an ethical committee was required. Therefore, only informed consent forms were used.

Interaction Design

In two studies, requirements and design guidelines for the interface for an advisory system were investigated. Specifically, it was investigated which type of advice should be provided and which modalities such an interface should employ.

In one study, it was investigated whether the advice should provide information about the desired acceleration or the desired speed target. With automated cooperative driving, the control system continuously adjusts the acceleration, complying with the insight that many small adjustments ensure a smoother traffic flow than fewer large adjustments. If we want advisory cooperative driving to mimic this behavior as much as possible, it implies that the system will present frequent advices for small changes in acceleration. However, we believe that the mental model of the driver is cast more in terms of speed targets, and that acceleration is only a low-level control parameter to reach a certain speed target. To investigate whether drivers prefer advice in terms of acceleration or speed targets, in one condition, participants received the acceleration advice calculated by an algorithm for automated cooperative driving, while in the other condition the acceleration advice was mapped onto a speed target advice. The visual displays for the two interfaces are shown in Figure 2. Both subjective and objective measurements (speed response) were collected. The experiment was conducted with 29 participants.

Figure 1. Driving simulator setup. The driving scenario is shown on an array of five LCD screens (of which only the middle three are shown). A sixth screen below the middle screen shows the standard dashboard display with speedometer and RPM display. A small screen (18 cm diagonal), mimicking a display in the mid console, is attached to the right of the sixth screen to present the displays under investigation.

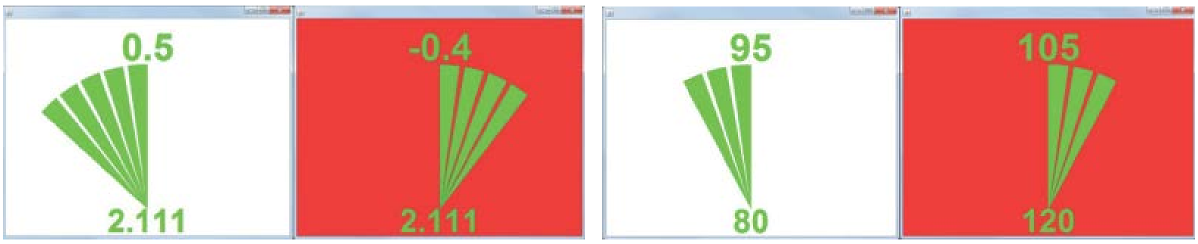


Figure 2. Left: Display for acceleration advice. Right: Display for speed advice.

It was found that eleven participants preferred the acceleration advice and eighteen preferred the speed advice. The ones preferring the speed advice said that the acceleration display allowed more precise control of the vehicle, while the ones preferring the speed advice said that the speed display allowed more freedom in how to implement the advice. The preference was also reflected in the performance: Those preferring the acceleration advice drove with a smaller time gap with the acceleration advice than with the speed advice, while for those preferring the speed advice it was the other way around.

In another study, it was investigated how to map the advice onto the visual and auditory modality. For advisory cooperative driving to provide the expected benefits on throughout, drivers should execute the advice as quickly as possible, turning the advice into an urgent message. Because of the inherent eyes-busy driving context, urgent messages are best communicated through the auditory modality, but at the same time the acceptance of auditory messages is an issue because of their obtrusiveness. In addition, the bandwidth for non-speech audio is limited, meaning that it may be hard to find auditory warning signals for different driver assistance systems that are sufficiently distinctive to be easily recognized. For this study, only speed advice was considered. In an iterative process, different concepts were explored and evaluated with experts. The final concept emerging from this process consisted of a multimodal interface in which a glanceable visual display informed drivers whether they were driving at an appropriate speed or too fast or too slow (Fig. 3), and auditory warning signals informed drivers how much to adjust their speed and when to stop speeding up or slowing down.

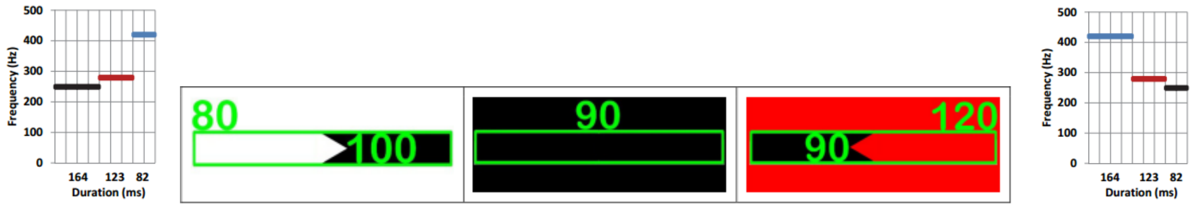


Figure 3. From left to right: Auditory warning signal for Speed up; Visual displays for Too slow, Appropriate, Too fast; Auditory warning signal for Slow down. The durations and inter-signal intervals of the auditory warning signals were adjusted to signal the amount of acceleration/deceleration needed.

To evaluate the concept with target users, several studies were set up with a driving simulator. In a study with 24 participants, the speed advice was presented through the auditory and visual modality, and participants' opinions concerning the usefulness, pleasantness, annoyance and so forth were elicited about different design options. In a study with 12 participants, different design options for the auditory warning signals were evaluated in terms of experience attributes (annoyance, understandability, mental load) and their effects on behaviour. For the latter, a metric was developed call Speed Response, which is the acceleration behaviour in a 5-second interval after the warning signal is presented. Typical graphs showing results are shown in Figure 4.

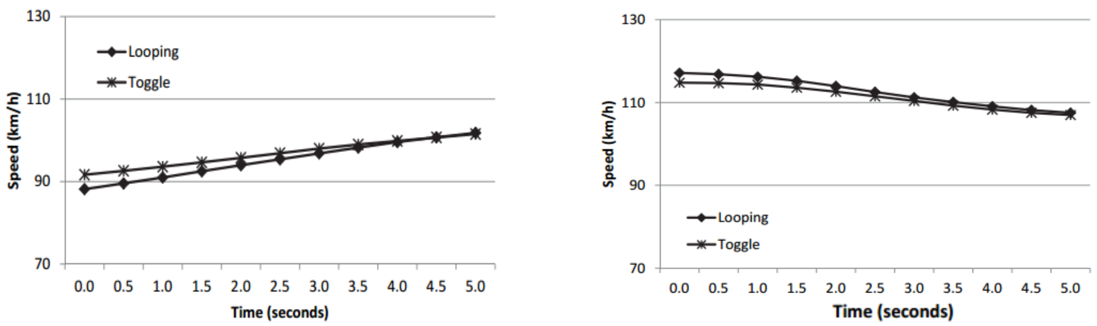


Figure 4. Speed response for two different auditory warning signals. Left: response to "speed up" warning signal. Right: response to "slow down" warning signal.

Enhancing the compliance of drivers with cooperative speed advice

A third study focused on the willingness of drivers to engage in cooperative driving. While cooperative driving has clear societal benefits, the benefits for individual drivers may not be immediately obvious, therewith potentially reducing the willingness of drivers to engage in cooperative driving. It was investigated how drivers could be convinced to engage in cooperative driving by providing them with persuasive messages tuned to their individual persuasion profiles. Persuasion profiles represent the idea that users differ in their sensitivity to (or tendency to comply with) messages representing specific persuasion strategies. The notion of persuasion profiles was already validated in other studies [3]. Six different persuasion strategies were designed for the driving context, drawing the driver's attention to the beneficial effects of cooperative driving in terms of driving safely, in a relaxed or sustainable way, saving time or money (because of avoiding speed fines), or experiencing enjoyment. Participants' profiles were determined through a questionnaire [4] and it was investigated whether their cooperative driving behaviour showed differential effects of messages that were compatible or incompatible with their persuasion profile. An experiment was set up comprising different sessions where 28 participants went through a baseline condition, a condition where they received only speed advice, and a condition where they received speed advice combined with a persuasive message. The speed advice was offered such that the obvious way to comply was by lowering speed and joining a platoon in the right lane, and non-compliance was signaled by maintaining or increasing speed, staying in or moving to the left lane and taking over the platoon. Again, subjective and objective measures were collected. The objective measures consisted of lane choice and speed response. It was found that speed advice itself already resulted in behavioural change and that persuasive messages did not enhance the behavioural change. No evidence was found that messages which were compatible with participants' persuasion profiles led to more compliance with the advice than incompatible messages.

Conclusion and Discussion

Three studies were summarized investigating how to support drivers in cooperative driving, using a driving simulator setup. Two studies focused on the design of the interface, investigating whether to present acceleration or speed advice and how to map the advice onto the visual and auditory modality. The third study focused on the question of how to raise the adoption of cooperative driving . With respect to the latter question, the hypothesis that messages that are compatible with drivers' persuasion profiles increase their willingness to engage in cooperative driving could not be confirmed. Of course, before taking this as refuting evidence for the hypothesis, more studies should be conducted, as the outcomes of the current study may also arise from the specific characteristics of the experiment. In this respect, several elements of simulator studies may cast doubt on their suitability to study behavioural effects of novel technologies other than cognitive aspects such as driver's understanding of the interface and distraction. In the first place, experimental studies involve volunteers, which may give a biased view on the behavioural effects (of course, field studies involving volunteers may suffer from the same bias). Secondly, with respect to behavioural effects of persuasive technology, what should matter in the end is the long-term effect; that is, we are interested to find out whether people adopt certain technology and how this technology affects their behaviour in the long term. In order to address this issue, the current study had people join for three sessions of one hour each, but one may question whether this is sufficient to study long term effects. In addition, in order to motivate the participants to stay involved in the experiment and conceal the purpose of the experiment, a game approach was taken, turning the experiment into a kind of game, which may have induced a different type of behaviour than what we aimed to study. Thirdly, in order to study behavioural effects in an experiment, we need to define both the experimental and control conditions, and the particular conditions may induce situation-specific strategies, turning the outcomes of the experiment into artefacts which are peculiar to the experiment and not indicative of the behaviour that would occur under realistic conditions. In sum, simulator studies may be perfectly suited for studying cognitive aspects such as usability and distraction, but less suited for studying behavioural effects relating to the adoption of the technology. For the latter, insights from simulator studies can only be preliminary and further evidence needs to be obtained through field studies, where people try to pursue their own goals instead of those imposed by the experimenter.

References

[1] van Arem, B., van Driel, C. J. G., Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems, **7**(4)*, 429–436.

[2] Van den Broek, T.H.A.; Ploeg, J.; Netten, B.D. (2011). Advisory and autonomous cooperative driving systems. In *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE)*, p. 179-180. January, 9-12, Las Vegas, NV, USA. IEEE.

[3] Kaptein, M., De Ruyter, B., Markopoulos, P., Aarts, E. (2012) Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. *ACM Transactions on Interactive Intelligent Systems, **2**(2)*, Article No. 10.

[4] Shahab, Q., Terken J., Eggen, J. (2013). Development of a questionnaire for identifying driver's personal values in driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, p. 202-208. New York, NY, USA: ACM.

Automation of training and assessment with DATA Centred Design (DCD)

Jorrit Kuipers

Green Dino BV, Wageningen
Delft University of Technology – 3ME, Delft, The Netherlands,
jorrit@greendino.nl

Introduction

Operators of complex systems not only have to follow procedures but also have to recognize changes in the system and anticipate. Car drivers have to control their vehicle and participate in traffic at the same time. Traffic itself is a complex dynamic system with interaction between the operator, other traffic participants and traffic regulation systems. Novice operators make more mistakes due to the complexity of the system and have a high failure risk. All operators suffer from so called erosion of skills (EOS) in situations where automated behaviour is not stimulated enough. Ecological Interface Design (EID) is a worldwide accepted design methodology for interfaces of complex real time dynamic systems. Operator failures due to erosion of skills are not within the scope of EID. EOS can result in severe system failures. This problem can also occur with the automation of driving tasks in smart vehicles solutions. In this presentation I propose a new method for interface design that incorporates EOS.

Erosion of skills – EOS -

The erosion of skills due to computer based automation is a major issue in aviation. Many plane crashes are directly related to EOS. Antonovich [1] reported that flight crews loose situational awareness as a result of automation. Other reported concerns include: loss of manual skill, overconfidence in automation, and difficulty in predicting and monitoring what the automation is doing or will do. These concerns are related to erosion of skills and pose a significant threat to air traffic safety.

Antonovich reported that the decrease of instrumental feedback due to automation is one of the factors behind EOS. Olson [2] stated that "in the absence of salient indications (i.e., Dashing lights, colour changes, etc.), pilots often do not pay attention to potentially relevant information." Without salient information the crew is likely to overlook or miss important or critical information.

EOS due to computer automation is widely known as “out of the loop syndrome.”. The real cause however is not computer automation but brain automation. The automation of behaviour is a process in the brain that helps us to process the enormous load of information without much mental effort. The repetition of stimuli associated with a stored reaction (behaviour) strengthens the nerve connections that form patterns (information clusters) in the brain. In contrast to little stimulation can weaken a brain pattern. When a particular brain pattern is too weak, the reaction to the stimuli will not be activated. The behaviour seems to be forgotten. Because we are not aware of this process, the skill can erode unnoticed. The frequency and strength of stimuli have a direct influence on the process of strengthening or weakening of brain patterns. This confirms the observation of Olson about the absence of salient indicators and the observed performance drop of flight crews.

The relationship between performance and information load is described in the so called Yerkes–Dodson law [3]. Yerkes-Dodson’s law incorporates the influence of stress and arousal, on performance. In high and low levels of information load the performance drops. Low levels of information can lead to boredom. When salient stimuli are missing, EOS starts.

Erosion of skills is not only a specific problem for operators who have to deal with computer automation, but it is a generic process related to the working of the brain. Automation of behaviour and skill acquisition, are the result of stimuli repetition. A high frequency of stimuli (that exceed the action potential threshold) increases automation but when the frequency drops the erosion of skills starts. Staying “in the loop” will help to stabilise or increase performance of (novice) operators during and after training and in the case of computer automation.

Ecological interface design – EID -

Jens Rasmussen [4] and Kim Vicente [5] developed a method for interface design, Ecological Interface Design (EID), aimed at improving the performance of operators and at the same time increase system stability. The term ecological originates from ‘ecological psychology’. This area of psychology focuses on relations between mankind and his natural environment. EID differs from ‘user centred design’ (UCD), which focuses only on the operator. With UCD the emphasis lies on the operator of the interface and his experience executing a specific task. EID goes further and projects the operator’s experiences in the operation of the total system. A system oriented approach instead of an operator oriented approach.

Ecological Interface Design is a structured method that keeps the mental workload low to improve the learning process and prevent operator failure as the result of an overload of working memory. EID stimulates knowledge based reasoning. Knowledge based reasoning supports the operator in unknown situations in which he cannot rely on automated behaviour. Low workload is necessary for knowledge based reasoning because the mental effort required is high. Automation of tasks supports the improvement of operator performance but it can also result in erosion of skills. EOS occurs in case the operator gets insufficient (salient) feedback from the system to keep automated skills at the desired performance level. EID should be improved to take account of EOS.

Data Centered Design – DCD -

EOS is an unconscious process. Therefore it is difficult for the operator to take precautions. External feedback is necessary to prevent operators from EOS. Accompanied driving for novice drivers is a good example of how to prevent younger car drivers for EOS after they get their driver’s license. Parents deliver the necessary strong human feedback the instructor gave during driving lessons. After many kilometers the younger drivers learn how to interpret the weaker feedback from the traffic system. It takes about 5.000 kilometers to decrease accident risk by 50%.

EOS can be monitored by continual observation, data recording and data analysis. Data analysis is useful for pattern recognition. This can support the operator in the process of reasoning to find solutions for unexpected situations. This results in reduced load on working memory. Kirschner et al. [6] emphasizes that knowledge based reasoning is not an effective method to solve problems.

Data recording and data analysis are not components of EID. Therefore the author proposes a new method for interface design, Data Centered Design (DCD), a data oriented approach. The data from the system is the main objective and not the description of the system like EID has. System relations are often to complex to describe and analyze with human observations. Data analysis provides the most insight into the changes in components of the system. Relations between events are found more easily than with knowledge based reasoning.

EID enhances the performance of system operators through automation of skills, keeping the workload low. DCD enhances the performance of system operators through monitoring EOS. Avoiding EOS is more important than automation of skills. DCD encloses monitoring of workload and automation of skills, like EID.

Virtual assistant

I used the DCD methodology to develop a virtual driving assistant for the automation of training and assessment of novice car drivers. The main target was to teach students the basic driving principles on a driving simulator without the help of a driving instructor. In order to accomplish this a didactical model called 'dynamic balance' was developed. The basic principle of this model is the automation of driving tasks with respect to the limitations of the short-term (working) memory, the capacity of the long-term memory and the erosion of skills. Feedback mechanisms have a central role in this model. Strong (salient) feedback makes the operator dependent and weak feedback increases the possibility of EOS. Continuous observation of the students’ performance on driving tasks (successes or failures) gives insight into the automation and erosion of skills. With help of real time adaptive instruction the level of feedback increases or decreases during training. The ‘virtual instructor’ called Victor, uses 3 adaptive instruction levels:

1. Do on instruction: the student gets precise instruction on every step in a driving task procedure.
2. Do with less instruction: the student gets indirect instruction. The driving task is introduced but not explained in detail.
3. Do without instruction: the student is only corrected.

Real time assessment on success or failure of driving task steps is necessary for adaptive instruction. The amount of feedback is based on the performance level of the student. Too much (overload), or too few (underload), instructions will slow down automation of behaviour. Separation of driving tasks is also necessary to prevent overload of working memory. A new driving task can only be added to the training in case the active task is automated. In other words, the student performs the task several times successfully without any feedback to prevent to high workload and a slowing down of the learning process.

In the driving simulator administration system the achieved instruction levels are reported per lesson (Figure 1). Quick students can achieve level 3 on driving task within a driving lessons. An average student will sometimes need 2 lessons. When a student stays at level 1 the lesson should be rehearsed. The student is not ready for a new driving task. A failure on a driving task forces the instruction to a lower level. In case the student fails one time in a row, the full instruction is given once at level 1. The level remains at 3. Two or more mistakes in sequence will lower the instruction level permanently. The automated skill has eroded. In the student report (Figure 1) this is the case for the driving task 'maximum speed' and some 'giving right of way' tasks. After a set of successes in sequence, the instruction level goes up again.

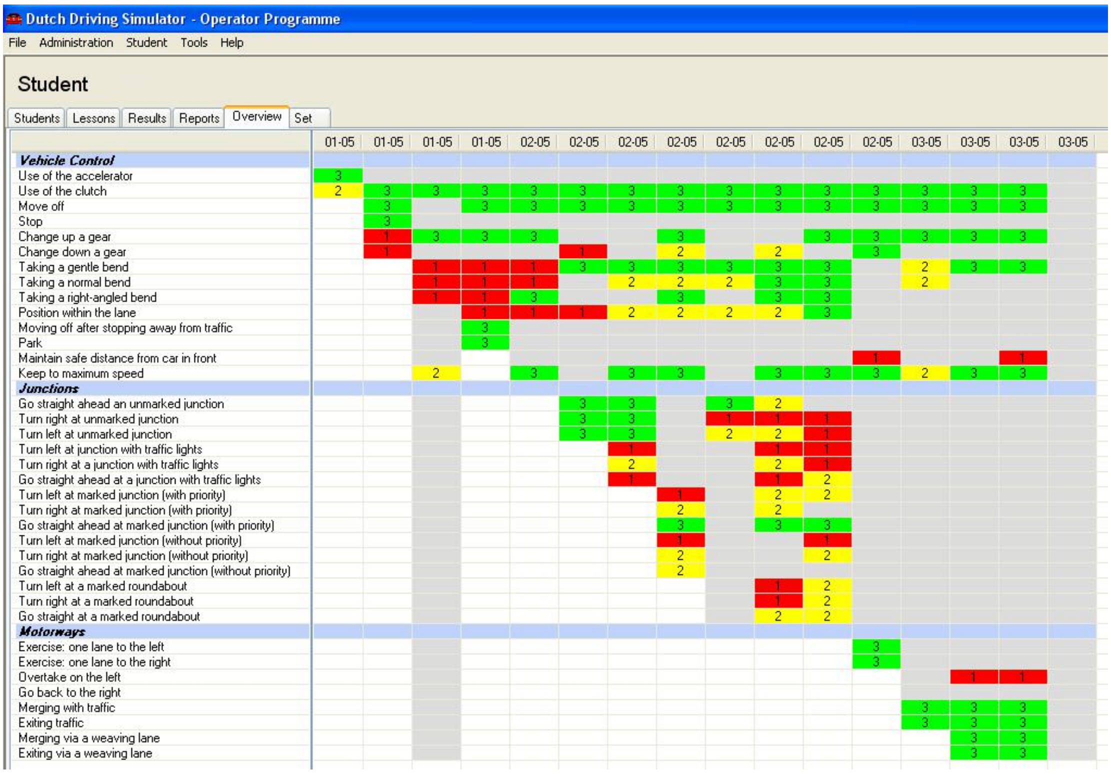


Figure 1. Screenshot of the Administration System ‘Overview: instruction levels’.

De Winter [7] compared the performance on the first examination between driving simulator students and those who only followed car lessons. The driving simulator students got instruction from the virtual driving assistant. The graduation percentages of 804 simulator students (reference group N=1774) were compared with the average national graduation percentage in 2005.

The results showed a 4% to 5% higher passing rate of the driving simulator students on their first examination than the national average. The scores are not corrected for self-selection, however driving schools with a simulator have a relative high number of students and tend to the average. A regression analysis showed a relatively strong negative association between the number of simulator training blocks and the duration of driver training. These results indicate that driving lessons with the virtual driving assistant are an effective alternative for on road driving lessons.

The interface design methodology DCD seems to be effective for the training of novice car drivers. However no research been conducted to this point measuring the effects of simulator instructed driving, on road safety. After passing the examination the novice car drivers where no longer observed. They didn’t get feedback on eroded skills anymore. I suppose that unless DCD methodology is also used after licensing, EOS will remain a major cause of traffic accidents.

Acknowledgment

This research is supported by the Dutch Ministry of Economic Affairs and Green Dino BV.

References

[1] Antonovich, B. (2008). The flight crew and automation. *The Journal of Aviation/Aerospace Education and Research, **17(3)***, Article No. 3.

[2] Olson, W.A. (2001). *Identifying and mitigating the risks of cockpit automation*. Wright Flyer Paper No. 14.

[3] Yerkes, R.M., Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, ***18***, 459-482.

[4] Rasmussen, J. (1983). Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, ***13***, 257-266.

[5] Vicente, K.J. (1999). Ecological Interface Design: Supporting operator adaptation, continuous learning, distributed collaborative work. In *Proceedings of the Human Centered Processes Conference*, p. 93-97.

[6] Kirschner, P.A., Sweller, J., and Clark, R.E. (2006) Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, **41**(2)*, 75–86.

[7] De Winter, J.C.F. S., De Groot, Mulder, M., Wieringa, P.A., Dankelman, J., Mulder, J.A. (2009). Relationships between driving simulator performance and driving test results. *Ergonomics*, ***52***, 137–153.

Using a virtual grocery store to simulate shopping behaviour

E. van Herpen[1], T. Yu[2,3], E. van den Broek[1], H.C.M. van Trijp[1]

[1] Marketing and Consumer Behaviour Group, Wageningen University, Wageningen, the Netherlands
{erica.vanherpen, e.vandenbroek, hans.vantrijp}@wur.nl
[2] Noldus Information Technology, Wageningen, the Netherlands
[3] Essensor, Ede, the Netherlands
tianna.yue@gmail.com

Introduction

Simulating a store environment by using virtual reality techniques offers important potential advantages for research into consumer behaviour. Through the use of virtual reality, the store can be simulated in a realistic and cost-efficient way [1,9]. This allows researchers to collect data in a tightly controlled but realistic store environment, at relatively low cost and high flexibility [3]. Virtual supermarket systems are being used to study consumers' reactions to price changes of food products [12], emotional responses to retail environments [6], and responses to emptied shelf space [4].

In order to fulfil its promise, a good understanding of which types of responses in a virtual environment resemble real-life behaviour, and which do not, is vital. Identifying areas where caution is needed can prevent unjustified generalizations of virtual reality outcomes [3]. At the same time, it is important to understand what the added realism of virtual reality has to offer, over and above the use of simpler pictorial stimulus materials. For instance, prior research has shown that an increase in visual realism can enhance spatial learning of a virtual layout [7] and effectiveness and efficiency in navigation tasks [8]. The present study builds on and extends this work by comparing a choice task using virtual reality to both a shopping trip in a real brick-and-mortar supermarket (with a similar choice task) and a choice task using product pictures. The key question is in which ways the choice behaviour of consumers is well represented in a virtual supermarket, and on which variables there are deviations from reality.

A few prior studies have examined how consumer choices in a virtual store environment compare to actual sales data (e.g., [2,3]). Yet, there are many reasons why market shares may differ from choices made in a virtual environment, and not all are related to the use of virtual reality as such. Benchmarking against existing methodologies is important to gain insight into the relative strengths and weaknesses of these methods. In virtual reality, as opposed to a pictorial representation, people have a better view of individual products from all angles, have a better feel for the shelf space assigned to products and the way shelves are organized, can manoeuvre through the store, and get a better feel for the overall store atmosphere. Taken together, this should increase people's sense of being present in the virtual store and possibly allow for more habitual consumer behaviour to occur.

The virtual supermarket

The virtual supermarket was displayed using a PC with three LCD screens of 42 inch each, which resulted in a 180 degree field-of-view. The shopping simulation software was developed in collaboration with Green Dino BV (www.greendino.nl), and the research conducted was part of the FOCOM project (http://www.focom-project.net/). Participants could navigate through the virtual supermarket using keyboard and mouse. Eye-level was set according to the average adult height. Participants could examine a product in more detail by double clicking on it when standing in its vicinity. This provided them with an enlarged front view of the product itself, and an enlarged view of the shelf tag, which contained a description of the product and its price. The software kept track of the products chosen, the time at which products were selected, and the time that participants spent scrutinizing specific products.

Pretest

Before conducting the main experiment, we performed a pretest (N = 90). The objective of this pretest was to examine whether a virtual supermarket environment could engage people to feel more present in the store than a more traditional approach (i.e., using product pictures) would. A sense of "presence" is often considered key to virtual reality experiences [5,10,11,13]. It implies that a person has a sense of being in the virtual environment (rather than being in the research lab). The pretest compared a mock shelf display, a representation of this shelf display in virtual reality and a picture of the same shelf display. Results indicate that the use of a virtual environment can indeed increase the perceived level of presence compared to a pictorial display.

Main experiment

The objective of the main experiment is to examine how consumer behaviour in a virtual environment compares to behaviour in a real-life store and to behaviour when pictorial stimuli are used. Participants (N = 100) were recruited by telephone from a consumer panel, and randomly assigned to one of three conditions in a three group design. The pictorial and virtual reality conditions were constructed to match the assortments of fruit/vegetables, biscuits, and milk of a local supermarket. Data collection in the real life store took place one week prior to the data collection using representations of this store, in order to allow for the pictorial and virtual reality representations to be matched closely to the real life situation (e.g., sales promotion material). Examples of the biscuits aisle are provided in Figure 1.

Data collection in the physical store condition took place in a local supermarket. This supermarket contained 31 milk products (none of which had a promotional offering during the data collection period), 32 fruit/vegetable products (five of which were on sale), and 230 biscuits (nine of which were on sale during the week of data collection). In this condition, participants were greeted by a first research assistant at the entrance of the supermarket. They received a brief task instruction and a shopping list containing the three focal product categories. The procedure in the other two conditions was as similar as possible, with the exception that participants in the virtual reality condition could first practice in a virtual store for as long as they needed to feel comfortable with the setup. Participants received the same instruction and shopping list.The virtual supermarket showed a realistic simulation of the actual store. All products from the product categories were presented, in the same shelf position as in the actual supermarket and participants could walk around in the virtual environment using a predefined shopping route and scrutinize products by using the keyboard and mouse.The picture condition showed static screenshots of the virtual store on a regular desktop screen (19 inch). Screenshots showed (a) a view of the whole virtual store, (b) an overall view of the target aisles, (c) a closer view of the shelves with index numbers assigned to the products, and (d) pictures of the each individual product as well as its shelf tag. These were connected using hyperlinks, so that participants could switch from product to product, or from shelf to shelf.

Key dependent variables in this study are the amount of products selected from each product category, the amount of different products selected (as a measure of variety seeking), average price paid, proportion of products bought on sale, and proportion of products bought from top/middle/bottom shelves.

Figure 1. The biscuits aisle in reality, virtual reality, and as picture on screen.

Results and discussion

ANOVA analyses were used to examine the amount of products bought for each product category, across the conditions. Results show that consumers tend to buy more products in lab conditions than in the actual store. Examining the amount of different types of products bought (a measure of variety) shows that the increased purchases in the virtual environment appear primarily due to increased variety seeking. Moreover, for several behavioural measures, the behaviour in the physical store is more closely resembled in the virtual environment than in the picture condition. For instance, in the milk category, the virtual supermarket was better able to simulate responses to vertical shelf position. Taken together, results indicate that although not all potential biases in behavioural responses disappear, the virtual supermarket has advantages over the use of pictures when attempting to simulate shopping behaviour.

This has important implications for the potential use of the virtual supermarket by retailers and manufacturers. A virtual store can be used to study a large variety of relevant issues regarding consumer food choices, at the product level (e.g., packaging, labelling), the assortment level (e.g., shelf space allocation) and the store level (e.g., store layout). The current study indicates that a virtual store can provide insightful information on consumer behaviour, and can more accurately represent this behaviour than is accomplished with pictorial stimuli.

References

[1] Berneburg, A. (2007). Interactive 3D simulations in measuring consumer preferences: Friend or foe to test results. *Journal of Interactive Advertising, **8**(1)*, 1-37.

[2] Burke, R.R., Harlam, B.A., Kahn, B.E., Lodish, L.M. (1992). Comparing dynamic consumer choice in real and computer-simulated environments. *Journal of Consumer Research, **19**(1)*, 71-82

[3] Campo, K., Gijsbrechts, E., Guerra, F. (1999). Computer simulated shopping experiments for analyzing dynamic purchasing patterns: validation and guidelines. *Journal of Empirical Generalisations in Marketing Science, 4(2)*, 22-61.

[4] van Herpen, E., Pieters, R., Zeelenberg, M. (2009). When demand accelerates demand: Trailing the bandwagon. *Journal of Consumer Psychology, **19**(3)*, 302-312.

[5] Lessiter, J., Freeman, J., Keogh, E., Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and virtual environments, **10**(3)*, 282-297.

[6] Massara, F., Liu, S.S., Melara, R.D. (2010). Adapting to a retail environment: Modeling consumer–environment interactions. *Journal of Business Research, **63**(7)*, 673-681.

[7] Meijer, F., Geudeke, B.L., van den Broek, E.L. (2009). Navigating through virtual environments: Visual realism improves spatial cognition. *CyberPsychology & Behavior, **12**(5)*, 517-521.

[8] Meng, F., Zhang, W., Yang, R. (2014). The development of a panorama manifestation virtual reality system for navigation and a usability comparison with a desktop system. *Behaviour & Information Technology, **33**(2)*, 133-143.

[9] Ruppert, B. (2011). New directions in the use of virtual reality for food shopping: marketing and education perspectives. *Journal of Diabetes Science and Technology, **5**(2)*, 315-8.

[10] Schuemie, M.J., van der Straaten, P., Krijn, M., van der Mast, C.A. (2001). Research on presence in virtual reality: A survey. *CyberPsychology & Behavior, **4**(2)*, 183-201.

[11] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, **42**(4)*, 73-93.

[12] Waterlander, W., Scarpa, M., Lentz, D., Steenhuis, I. (2011). The virtual supermarket: An innovative research tool to study consumer food purchasing behaviour. *BMC Public Health, **11***, Article No. 589.

[13] Witmer, B.G., Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments, **7**(3)*, 225-240.

Driver state estimation: from simulation to the real world

Tobias Heffelaar[1], Renske B. Landman[2,3], Menno Merts[3], Jasper Michiel van Hemert[4], Arjan Stuiver[5], Lucas Noldus[1]

[1] Noldus Information Technology BV, Wageningen, The Netherlands
t.heffelaar@noldus.nl, L.Noldus@noldus.nl
[2] ErgoS Human Factors Engineering, Enschede, The Netherlands
renske.landman@ergos.nl
[3] HAN University of Applied Sciences – Automotive, Arnhem, The Netherlands
menno.merts@han.nl
[4] TomTom, Amsterdam, The Netherlands
jaspermichiel.vanhemert@tomtom.com
[5] TNO, Soesterberg, The Netherlands
arjan.stuiver@tno.nl

Introduction

The task of driving has become more complicated over the years. Advanced Driver Assistance Systems (ADAS) and in-vehicle information systems (IVIS) have introduced computer technology in cars. Despite the aim of a significant number of these systems to support drivers they can also increase the complexity of driving. Combining this with the fact that distraction plays a role in most (near) accidents [1], it is crucial to ensure that newly introduced systems do not add complexity nor increase distraction.

It is common to test the impact of systems on driver performance, distraction and cognitive workload, as a means to ensure that new (ADAS) systems do not overload nor distract drivers [2]. A driving simulator creates a controlled environment in which it is possible, besides being able to provide a reproducible driving scenario, to experiment with situations that are very rare (e.g. flooding of roads) or that are simply too dangerous (e.g. near-accidents) to study in the real world.

Conducting experiments in a simulated environment, like a driving simulator, is only of value if the measured effects can be transferred to the real world. A number of experiments have shown that driver behavior on the road is comparable to driver behavior in a driving simulator [3 - 5]. Although the results are similar, behavior is not identical. A careful experimental design can optimize the comparability of the behavior of a driver in the real and simulated world and therefore help to make the transfer from simulation to the road [5].

ADVICE

The RAAK [6] project Advanced Driver Vehicle Interface in a Complex Environment (ADVICE [7]) focusses on the ability to assess driver state in real-time in the real world [8]. The project is a cooperation between HAN University of Applied Sciences – Automotive, Noldus Information Technology, TNO, TomTom and Delft University of Technology. Driver state has been defined as a combination of personal factors that affect driver performance. Examples of these factors are mental workload, fatigue, alertness, drowsiness, and driving skills. The Driver State Estimation (DSE) developed in ADVICE has been evaluated in a series of real-world driving experiments, using a BMW equipped with behavioral measurement systems by Noldus IT and the HAN University of Applied Sciences. In order to assess the performance of the DSE a combination of driving performance parameters (e.g. steering reversal rate, headway), and driver state parameters (e.g. heart rate variability, pupil dilation) were selected and tested in a driving simulator in a set of driving conditions specifically designed to match the real-world testing environment.

Figure 1. Experimental BMW 'in action'.

Figure 2. The Observer XT: integrated display of the driver's main front screen with gaze overlay, gaze events (red colored bars), driving tasks (multi-colored bars), and vehicle parameters (red/green/blue lines).

For the simulator experiments DriveLab was used. DriveLab is a novel, integrated system for the analysis of driver behavior, based on established proven technology from Green Dino, Noldus Information Technology and Smart Eye [9].

Based on the results of the simulator experiments the most sensitive parameters were selected for use in the real-world experiments. Also the experiments in the driving simulator and the real world have been designed to elicit identical driving behavior. This results in a set of parameters recorded, with the same equipment and in similar driving situations, in the real world and in the simulator. The results of both experiments will be compared and presented. Based on the results, a statement will be made on the suitability of the method of selecting real-world parameters based on simulator experiments.

References

[1] Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R. (2006). *The 100-car naturalistic driving study: Phase II – Results of the 100-car field experiment.* National Highway Traffic Safety Administration, U.S. Department of Transportation. Report No. DOT HS 810 593.

[2] Chisholm, S.L., Caird, J.K., Lockhart, J., Fern, L., Teteris, E. (2007). Driving performance while engaged in MP-3 player interaction: Effects of practice and task difficulty on PRT and eye movements. In Proceedings of *Driving Assessment 2007: 4th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Article No. 41. July 9-12, Stevenson, WA, USA. Iowa City, IA, USA: The University of Iowa.

[3] Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Akerstedt, T., Bioulac, B. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep*, ***28****(12)*, 1511-1516.

[4] Blane, E. (1996). *Driving simulator validation studies: A literature review*. Institute of Transport Studies, University of Leeds. Working Paper 480.

[5] Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L.A., Coughlin, J.F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, ***53****(3)*, 404-420.

[6] RAAK: Regional Attention and Action for Knowledge circulation, http://www.innovatie-alliantie.nl/english.html [Last accessed on January 27, 2015].

[7] Advanced Driver Vehicle Interface in a Complex Environment (ADVICE) project: http://www.noldus.com/projects/advice [Last accessed on January 27, 2015].

[8] Alders, M., van Hemert, J. M., Pauwelussen, J., Heffelaar, T., Happee, R., Pauwelussen, J. (2012). Managing driver workload using continuous driver workload assessment. In A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman (Eds.), *Proceedings of Measuring Behavior 2012: 8th International Conference on Methods and Techniques in Behavioral Research*, p. 38-39.

[9] Heffelaar, T., van Mil, B., Kuipers, J., Andersson, J., Wiertz, L., Noldus, L.P.J.J. (2014). DriveLab: understanding driving behavior made easy. In N. Stanton, S. Landry, G. Di Bucchianico, A. Vallicelli (Eds.), *Advances in human aspects of transportation, Part III*, p. 351-354. Danvers, MA, USA: AHFE Conference.

Measuring behaviour in a manoeuvring simulator

W.M. Uitterhoeve

MARIN, Wageningen, the Netherlands
w.uitterhoeve@marin.nl

Introduction

Several sources state that 70 to 80 percent of maritime accidents are caused by human error. Technical improvement in vessel design or operational equipment generally contributes to more safety on board and to a decrease of accidents in the maritime field. As a result, the contribution of human errors as one of the causes of maritime accidents relatively increases. The human factor is not negligible in complex manoeuvres.
To study the human factor in a maritime environment is possible in several ways. A manoeuvring simulator provides a quite realistic setting in which it is possible to study for example workload and performance in relation to task demand.

At MARIN any combination of two full mission bridges and four additional cubicles makes it possible to simulate nautical operations in applications like nautical research into new harbour and fairway layouts or the development (and training) of operational procedures. Human factor aspects like workload and performance measured during these simulator runs provide (objective) feedback to human behaviour.

Workload and performance measurements

The relation between workload, performance and task demand as mentioned by De Waard [2] is the basis for a methodology in which several parameters identify the performance and workload. The certainty of the results increases when more parameters point in the same direction.

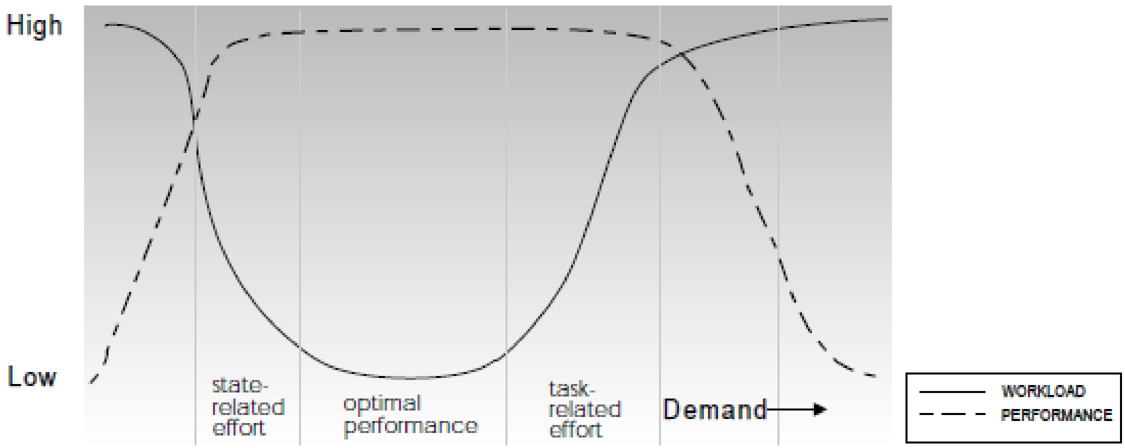


Figure 1. Workload and performance in six regions [2].

Subjective effort ratings: For subjective effort rating the RSME developed by Zijlstra [9] is used. This rating scale runs from 0 to 150 and contains levels from “absolutely no effort” to more than “extreme effort”. The candidate puts a mark on this scale.

Control equipment (Decon Medical Systems, Weesp, The Netherlands) records Objective heartbeat measurements. An elastic belt around the chest of the trainee contains a device to record an electrocardiogram and determines inter beat intervals (RR intervals). Via a wireless connection between the belt and an antenna connected to the measurement laptop the results are directly visible on the screen.

For on board measurements, when WiFi is not preferred, the Firstbeat Bodyguard is used. This device is attached to the skin with two chest electrodes and also determines RR intervals. With associated software it is possible to download locally stored data. The Bodyguard is designed also for long-term recordings (24h).

For both measurements, KUBIOS-software [4] is used to analyse RR interval time traces. From this Heart Rate Variability analysis both the time and frequency domain results are used in the interpretation of workload.

During the experiments, subjective performance rating is done by simulator instructors, who are experienced former seafarers and supervisors of the runs.

An objective performance judgement is based on independent simulator data. Predefined criteria for rudder usage, used power, swept path and speed for example help to quantify the controllability of the vessel during the manoeuvre.

A peripheral detection task measures focus on the main task. The secondary task applied in the experiments consists of reacting to a flash light in the peripheral view of the candidate, see Figure 3. The reaction time and missed stimuli are the indicators for focus on the main task [3].

Observations during the simulator runs by the researcher and recorded by video cameras provide behavioural information. For example someone standing at the bridge close to the displays and holding the tillers or VHF permanently shows another behaviour than someone standing back and overlooking the situation.

Technical or behavioural analysis boundaries

During the analysis of the results, simultaneous interpretation of all workload and performance indicators creates an overall picture of the situation. Statistical data of parts of the time trace are compared. Depending on the research questions the effect of events within one run are studied or several runs are compared to study the effect of changes in the scenarios.

Boundaries of analysis intervals are determined in several ways. When the focus is on the impact of a certain part of the manoeuvre on the operator, "technical" triggers (like the position at the fairway or the distance to an encountering vessel) mark the analysis boundaries. When more cognitive aspects matter, boundaries are related to the moment the operator perceives specific visual or auditory information. A combination of both gives insight into the effect of task demand or demanding circumstances on the duration and quality of the cognitive process before the operator takes action.

Practical examples

Since 2010 several experiments are executed at full mission bridge manoeuvring simulators, using above mentioned measurements. In the beginning the experiments focussed on the applicability of the methodology. Later on the measurements are applied to study questions related to workload. Some examples are described below.

Experiment 1: Effect of increasing task demand on workload [5]. In this study pilots trained a complex journey before they had to sail a new built vessel in real life. During the training the focus is to the more demanding parts of the journey. Based on his experience the instructor classifies the runs as easy, moderate or difficult. The results show that in the difficult runs performance decreases and workload increases, compared to the easy runs. After the training the measurements are also executed on board. For some candidates a direct comparison between the simulator and real life situation was possible. Especially in the difficult situations the results indicate a higher workload level as experienced in the simulator. The effect of longer working hours and the pressure of the real operations are expressed in this result.

Experiment 2: Application of workload and performance indicators in order to follow training progress [6]. During a Maritime Resource Management (MRM) training for Maritime Officers at the Maritime Simulation

Training Centre at Terschelling workload and performance indicators are related to the conscious competence learning curve [1], see Figure 2(left). A total of forty-eight students participated in the experiment. For only twenty-four students a combination of workload and performance measurements are available. The results show that seventeen out of twenty-four students made one or two steps within the learning curve, see Figure 2(right). Nine of them reached the end of the learning curve, characterised by a decrease in workload and increase in performance. The other eight benefited from additional training to master the new skill.

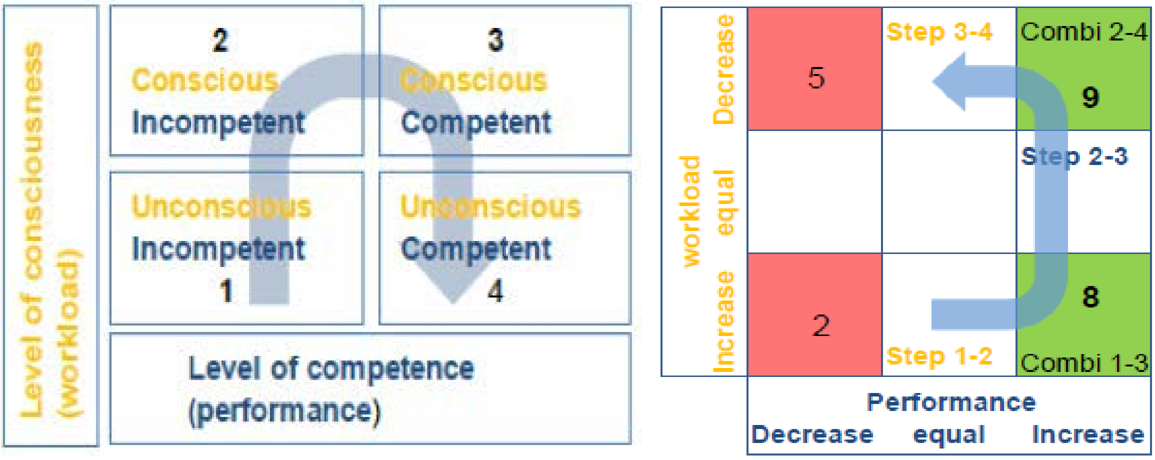


Figure 2. left: conscious competence learning curve related to workload and performance, right: amount of students per increased / decreased workload and performance.

Experiment 3: How differences in working strategy express workload differences for pilots and skippers [7]. While sailing at the same channel, changes in workload are studied for both pilots and skippers. Both experienced more workload in a condition without lights, but the amount and origin of the change differs. Due to differences in vessel and bridge design, skippers and pilots developed different working strategies. The pilots are more used to controlling the process and anticipate on future actions. On the other hand, skippers are busier with reacting to actual disruptions and correcting effects of past actions. For the skippers workload increases due to an accumulation of demanding factors, while for the pilots the main cause lays in the absence of reference lights in the environmental view.

The skipper mostly uses the radar for course keeping. He detects an encountering vessel at the radar, but has to complete the picture with information from the outside view. Due to inappropriate instruments for this inland condition, the pilot mainly uses the outside view for course keeping and collecting information. Event analysis demonstrated for example that during the non-lighted condition the skippers need more time to react to the second encountering vessel, while the pilots perform the same.

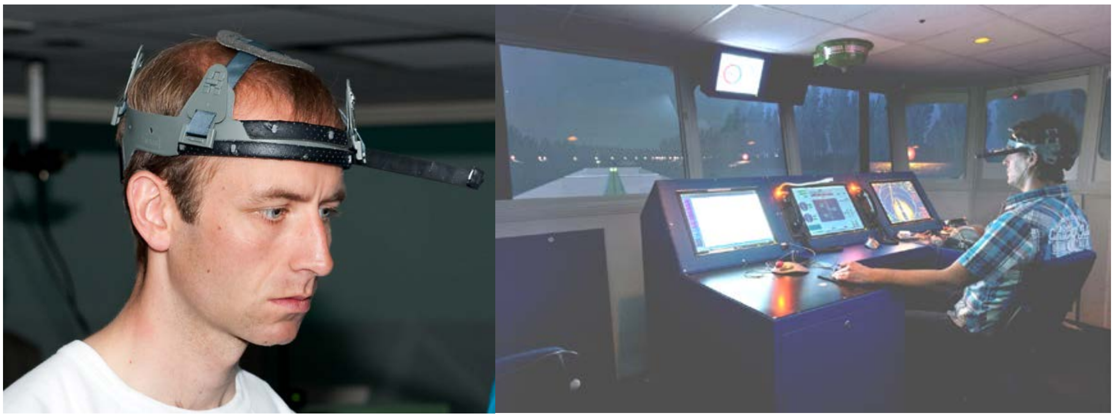


Figure 3. Candidate wears the secondary task helmet (left), impression of experimental setting with skippers (right).

Experiment 4: The effect of watch schedules on mooring master's workload during offloading operations [8]. This recent work is part of an ongoing thesis about offshore offloading operations which also includes a financial risk analysis. Mooring masters participated voluntarily in the experiments. HRV data and rating scales obtained during mooring master training in the simulator will be compared to on board data. The goal is to study the effect of watch keeping schedules on fatigue and performance and its relation to operational risks. The probability of an accident is calculated for one or two mooring masters on board, working in different schedules.

Connection with real-life and future additions

The pilot's contribution in experiment one increased their awareness of human failure, especially related to fatigue. During the next real operation they scheduled more recovery moments to increase alertness. Safety is an important factor in the offshore world. In experiment four, mooring masters are willing to voluntarily execute measurements during their work on board, driven by personal motivation and safety awareness.

Although, the technical simulator setting is utterly reliable (the modelled vessel behaviour in external environmental forces, bridge equipment, and outside view), more social and human factor related topics are missing in veracious simulations. For example experiment 1 demonstrated the differences in workload due to the operators being out of their daily activities and stress, working in the simulator during office hours and staying in a hotel. And in the offshore setting from experiment 4 the journey with associated jet lag at the start of the operation plays an important role.
When translating answers from a simulator setting to on board situations, it is a challenge to minimise the gap between simulator and on board settings and take into account the non-technical aspects. Collecting on board data and comparing these measurements with data from the simulator is essential in this.

To overcome this gap MARIN works towards a Maritime Human Factor Observatory: an environment to study behaviour of maritime personnel, using the above mentioned methodology extended with task registration, eye tracking and other tools, observing not only in an advanced simulator but also on board using mobile toolkits. Topics like fatigue during 24/7 operations, decision making, (individual) task optimisation, team work and communication can be studied in a controlled but realistic setting and translated into on board working limits.

Figure 4. The human factor observatory provides possibilities to study individual and team task optimisation.

References

[1] Burch, N. (1970s). *Learning a new skill is easier said than done*. http://www.gordontraining.com/free-workplace-articles/learning-a-new-skill-is-easier-said-than-done/. [Last accessed on March 27, 2014].

[2] de Waard, D. (1996). *The measurement of drivers' mental workload*. PhD-thesis, Traffic Research Centre, University of Groningen, Groningen, The Netherlands.

[3] Martens, M.H., van Winsum, W., (2000). Measuring distraction: The peripheral detection task. http://www-nrd.nhtsa.dot.gov/departments/Human%20Factors/driver-distraction/PDF/34.PDF [Last accessed on March 27, 2014].

[4] Tarvainen, M.P., Niskanen, J.-P. (2008). *Kubios HRV version 2.0, User's Guide*. Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Kuopio, Finland.

[5] Uitterhoeve, W., de Roos, R., Kluytenaar, P., Valk, P. (2012). Measuring workload and performance during complex manoeuvres, both on a manoeuvring simulator and on board. In *Proceedings of MARSIM 2012, International Conference on Marine Simulation and Ship Manoeuvrability*. Singapore, April 23-27.

[6] Uitterhoeve, W. *Workload Indicators measured and applied in order to follow training progress during an MRM training*. http://www.marin.nl/web/Publications/Publication-items/Workload-indicators-measured-and-applied-in-order-to-follow-training-progress-during-an-MRM-training.htm [Last accessed March 27, 2014].

[7] Uitterhoeve, W., Zeilstra, M. (2014). Differences in workload of both skippers and pilots due to changes in environmental bank lights. In N. Stanton, S. Landry, G. Di Bucchianico, A. Vallicelli (Eds.), *Advances in human aspects of transportation, Part III*, p. 73-85. Danvers, MA, USA: AHFE Conference.

[8] Van Vliet, K.J.A. (2014). *The effect of watch keeping schedules on mooring master's workload during offloading operations - A financial risk analysis of three watch keeping scenarios in FPSO - tanker offloadings*. MSc-thesis, Delft University of Technology. Delft University of Technology, Delft.

[9] Zijlstra, F.R.H., van Doorn, L. (1985). *The construction of a scale to measure perceived effort*. Department of Philosophy and Social Science, Delft University of Technology, Delft, The Netherlands.

Multi variable strategy reduces symptoms of simulator sickness

Jorrit Kuipers

Green Dino BV, Wageningen / Delft University of Technology – 3ME, Delft, The Netherlands
jorrit@greendino.nl

Introduction

Interactive 3D environments are valuable tools for the assessment and adjustment of human behavior. Driving a virtual car is cheaper, safer and more sustainable than driving on road [1]. However there is one major problem related to the use of 3D environments for assessment and training. A significant part of simulator operators suffer from symptoms resembling motion sickness, or more specifically simulator sickness. A common indicator for experienced operators is 30%, however this varies greatly. Symptoms can persist for long periods of time. Some drivers are not able to drive their vehicle after a simulator session [1]. This discomfort has a negative influence on the acceptance of 3D environments for assessment and adjustment of human behavior. Research by Johnson [2], Kingdon et al. [3] and Stanney et al. [4] indicates that in 1 to 1.5% of the operators that are exposed to a simulated environment, vomiting is induced

To prevent unnecessary delay of 3D environment development as a tool for the assessment and adjustment of human behavior, users and producers of interactive 3D environments will have to find effective methodology to reduce the severity of symptoms caused by simulator sickness. In addition new methods are needed to decrease the intensity of simulator sickness.

Variables

Differences in susceptibility for simulator sickness are not only observed between experienced and novice operators but also amongst experienced operators. Experienced car drivers react differently to specific configurations of driving simulators. They experience less symptoms of simulator sickness in a driving simulator with a wide view display, made with 5 LCD computer monitors (Figure 1), compared to a driving simulator with a wide view display made with 3 projectors (Figure 2). The main difference between these two simulators is the display surface. The total LCD monitor surface is $0.57m^2$ and the projector surface is $2.03m^2$. The display surface seems to be a variable that influences simulator sickness. A decrease in display surface results in lower susceptibility to simulator sickness. It is likely that there are more variables that have an influence on simulator sickness. Perhaps tuning those variables will reduce the severity of the symptoms and perhaps even solve the problem of simulator sickness.

Figure 1. Simulator with 5 LCD screens.

Figure 2. Simulator with 3 beamers.

In the literature many variables related to simulator sickness are discussed. A significant part of the available literature was produced by the US Army. The US Army did most of its research on simulator sickness in flight simulators and tank simulators [2][5] The variables can be classified as personal related, scenario related and technical related variables [6][7]. Table 1 gives an overview of variables related to simulator sickness. Variables with a strong influence on simulator sickness are: experience, history, session duration and optic flow.

Table 1. Simulator sickness variables.

Personal related variables		
Experience and age	Symptoms of simulator sickness can increase with experience. Becoming experienced is normally a process of many years. Therefore age and experience are correlated. Novice drivers experience less symptoms of driving simulator sickness than older drivers.	Miller et al. (1960) Reason et al. (1975) McGuiness et al. (1981) Crowly (1987) Hein (1993) Naoke et al. (2012)
Gender	Females seem more susceptible for simulator sickness than males. Males also need less time to adjust to a virtual environment like a driving simulator.	Hein (1993) Torant et al. (2000) Jeager et al. (2001)
History	A history of motion sickness, like car sickness or simulator sickness, correlates with the susceptibility for simulator sickness.	Gower et al. (1988) Lampton et al. (1994) Regan et al. (1994) Delahaye et al. (2004) Brooks et. al. (2010)
Scenario related variables		
Exposure	After a period of 5-10 minutes the symptoms of simulator sickness can grow fast.	Sparto et al. (1992) Jeager et al. (2001) Stanney et al. (2003) Bertin et al. (2007)
Velocity and acceleration	Navigation with high speed stimulates the symptoms of simulator sickness. Change of speed or direction stimulates the symptoms of simulator sickness.	Reason et al. (1975) Torant et al. (2000) So et al. (2001)
Scene complexity	Details needed for a realistic reproduction of a 3D	Van Cott (1990)

	environment stimulate the symptoms of simulator sickness.	Jeager et al. (2001) Kingdon et al. (2001) So et al. (2001)
Navigation	A restriction on free navigation stimulates the symptoms of simulator sickness.	Lackner (1992) Stanney et al. (1998) Jeager et al. (2001) Sharples (2008)
Technical related variables		
Motion	Motion systems with strong motion cues can increase the symptoms of simulator sickness. In particular when the motion cues and visual cues are not synchronized.	Kingdon et al. (2001) Johnson (2007) Plouzeau et al. (2013)
View angle and display surface	A large vertical and horizontal viewing angle stimulates the symptoms of simulator sickness. A large display surface stimulates the symptoms of simulator sickness.	Kennedy et al. (1989) Van Cott (1990) Sharples et al. (2008)

Sensory conflict

Several theories attempt to explain the mechanism of action that underlies simulator sickness. The neural mismatch theory of Reason [8] has been generally accepted. This theory relates simulator sickness to a sensory conflict during information processing which causes a mismatch signal in the brain. Divergent vestibular information plays a crucial role. People with a poorly functioning vestibular system do not suffer from symptoms of motion sickness. This mismatch signal stimulates the vomit center in the brain. As a reaction to this theory new methods that focus on increasing fidelity (realism) to solve the sensory conflict and decrease simulator sickness have been developed. Until now these methods haven't successfully dealt with the problem of motion sickness. I suggest a new approach. The mismatch signal seems to act like the neurotransmitter 'Substance P'. Substance P plays an important role in activation of the vomiting area in the brain. In the event of a mismatch signal causing symptoms of simulator sickness, weakening the signal can be a good strategy to decrease the severity of these symptoms. This particular method should decrease or even stop the release of the involved neural transmitter. The variables with the most influence on neurotransmitter release and activation of the vomiting centers of the brain should be identified and manipulated.

There are big differences in susceptibility of simulator sickness between operators of interactive 3D environments. Many variables are involved. The manipulation of specific variables has so far not led to an effective methodology for reducing simulator sickness. Pre-selection of operators based on motion sickness history seems a good method but is not effective enough and in many situations it is not desirable to exclude operators from a simulator session. A test run improves the effectiveness of preselecting. Preselecting can also help to determine which precautions should be taken to avoid simulator sickness.

Multi variable strategy

I propose a multi variable strategy for the alleviation of simulator sickness symptoms. Manipulation of multiple variables might be a good solution to increase the effectiveness of simulator sickness prevention. Variables with a high impact on simulator sickness are session duration and optic flow. For example session durations should not exceed 5 minutes. This strategy will keep the concentration of neurotransmitters below the threshold of activation of the vomiting centers in the brain.

The optic flow variable is influenced by variables like velocity, acceleration, angular acceleration and screen size. The screen size can be changed easily for driving simulators without effects for the scenario. Adjustments made to the vertical field of view (FOV) decreasing the image surface enormously. The display of graphics with wide screen monitors instead of computer projectors or adding black areas can decrease the image size up to 80% (Figure 3).

Figure 3. the vertical FOV is minimized with wide screens monitors and black areas.

This strategy seems to have a significant effect on most operators of driving simulators. Results of two commercial driving simulator tests with 111 and 826 participants respectively showed that this multi variable strategy decreased the percentage of operator who experienced motion sickness from 29% to less than 0.2%. The first is the base line test. Operators drove a standardized driving style test. 29% operators reported severe symptoms of motion sickness like drowsiness and nausea. This is a common rate for experienced drivers. The standardized driving style test was conducted using a driving simulator with beamer projection without using the multi variables method (Figure 2). The same driving style test was used in the second project. The multi variables strategy was used to lower the symptoms of motion sickness. The virtual FOV was decreased using the driving simulator with the LCD monitors (Figure 1) in combination with the black areas (Figure 3). In case the operator experienced motion sickness the black areas where activated. The session duration was maximized to 3 minutes followed by a break instead of 15 minutes continuous simulator use. 2% of the operators (15 persons) reported symptoms of motion sickness. After reduction of the vertical FOV with the black areas, only 0.2% (2 persons) reported still having related symptoms. One operator had to stop before the driving style ended. None of the operators displayed actual vomiting. The dropout rate of 0.1% is very low and a big improvement compared to single variable strategies like the motion sickness assessment questionnaire (MSAQ) of Brooks [9]. Books et al noted a 10% dropout.

Surveys held under driving simulator participants of the first and second test indicated not only a lower percentage of dropouts, but also an improvement of the user acceptance of interactive 3D environments as an alternative for a real environment. 24% compared to 16% in the first test.

Acknowledgment

This research is supported by the Dutch Ministry of Economic Affairs and Green Dino BV.

References

[1] de Winter, J.C.F. (2009). *Advancing simulation-based driver training*. Thesis. Delft University of Technology.

[2] Johnson, D. (2005). *Simulator Sickness Research Summary*. U.S. Army Research Institute for the Behavioral and Social Science.

[3] Kingdon K., Stanney K., Kennedy R. (2001). Extreme Responses to Virtual Environment Exposure. *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting, Vol. 45*, p. 1906-1910.

[4] Stanney, K.M., Hale, K.S., Nahmens, I., Kennedy, R.S., (2003). What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors, 45(3)*, 504-520.

[5] Johnson, D. (2007). *Simulator Sickness During Emergency Procedures. Training in a Helicopter Simulator: Age, Flight Experience, and Amount Learned.* Technical Report 1211. U.S. Army Research Institute.

[6] Kolasinski, E.M. (1995). *Simulator sickness in Virtual Environments*. U.S. Army Research.

[7] Cobb, S., Nichols, S., Wilson, J.R. (1995) Health and Safety Implications of Virtual Reality: In search of an experimental methodology. In *Proceedings of FIVE'95*, p 227-242.

[8] Reason, J.T., Brand, J.J. (1975). *Motion Sickness*. London, UK: Academic Press.

[9] Brooks, J.O., Goodenough, R.R., Crisler, M.C., Klein, N.D., Alley, R.L., Koon, B.L., et al. (2010). Simulator sickness during driving simulation studies. *Accident Analysis & Prevention, **42**(3)*, 788-796.

General Sessions

Automatic Statistics Extraction for Amateur Soccer Videos

Jan C. van Gemert[1], John G.M Schavemaker[2], Koen Bonenkamp[1,2]

[1]University of Amsterdam
[2]TNO

Abstract

Amateur soccer statistics have interesting applications such as providing insights to improve team performance, individual coaching, monitoring team progress and personal or team entertainment. Professional soccer statistics are extracted with labor intensive expensive manual effort which is not realistic for amateur matches. In this paper we develop a solution that automatically extracts action-related soccer statistics from a static camera pointed at the pitch. We implement a solution to player localization and action classification problem in human action recognition. Our method does not rely on player tracking, sliding windows, super voxels or construction of multiple hypotheses. Our work is developed with actual application in mind and a fully functional recognition pipeline is implemented, specifically tailored to meet the inherent challenges of action-rich soccer video.

Introduction

Professional soccer matches are manually annotated by people watching matches. From these annotated matches statistics are extracted which are used to inform the public with match overviews and fun facts. Bookmakers use these statistics to determine the likelihood of a match outcome. Team managers use statistics of opponents to develop strategies as well as statistics of their own team to improve teamwork and individual players. The majority of such statistics is extracted from video manually and is thus time consuming and expensive. For professional games the value of the data justifies manual labeling of every match. For amateur soccer matches, in contrast, the manual annotation effort is neither profitable nor possible given the huge amount of amateur games for children, adults and veterans. In this paper we propose an automatic method for statistics extraction from amateur soccer matches by placing a fixed inexpensive camera on the side of the pitch. From this video data, our method is able to automatically localize players and recognized the action these players are performing.

Related Work

Dense trajectories Wang et al. [9] reason that the 2D spatial domain and 1D temporal domain in videos show different characteristics and that it is more intuitive to handle them in a different manner than via interest point detection in a joint 3D space [1,2,3]. Wang et al. [9] densely track interest points through video sequences using optical flow. They (re)introduce the motion boundary histogram descriptor [1] that offers invariance in the case of camera motion. Dense trajectories in combination with motion boundary histograms consistently perform well in comparative studies [1,9]. For our own application we closely follow the approach by Wang et al. [9], our work differs however in that we also treat localization of actions.

Localization-by-detection Recent efforts in recognizing actions from longer video sequences evaluate an action classifier at various segments of the video [4]. Such an exhaustive search is computationally expensive because it requires scanning both the spatial and temporal location, as well as various spatial and temporal scales. This yields a huge 5-dimensional search space, or even a 6-dimensional search space if the aspect ratio of bounding boxes is not fixed. Several solutions have been developed that effectively avoid the exhaustive search, by only evaluating a subset of spatio-temporal windows (hypotheses) [2, 4, 5]. Jain et al. [5] generate a set of hypothesis bounding-boxes moving in time, such temporal bounding boxes are denoted *tubelets*. Derpanis et al. [2] implement an efficient template matching algorithm to detect actions. Action bank [6] builds on this template matching technique to create a high-level representation of video by constructing many individual action detectors. Our method differs from these works as we compute a single solution (hypothesis) directly without generating multiple hypotheses

Method

In Figure 1 we show an overview of our approach. For classification our work follows the approach by Wang et al. [9] closely, our work differs however in that we also treat localization of actions.

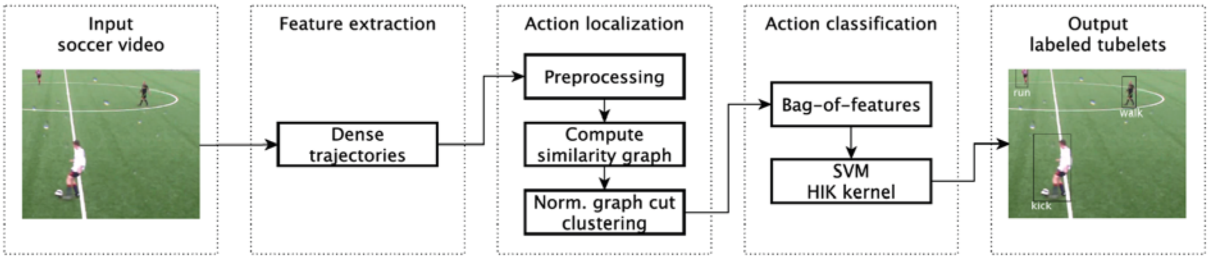


Figure 1. Action localization pipeline.

In action localization we cluster trajectories into groups that correspond to individual actions. In Figure 2 (left) we show raw trajectories. In Figure 2 (right) we show the ground truth obtained by manual player annotation of the video, the ground truth is the ideal clustering we aim to obtain.

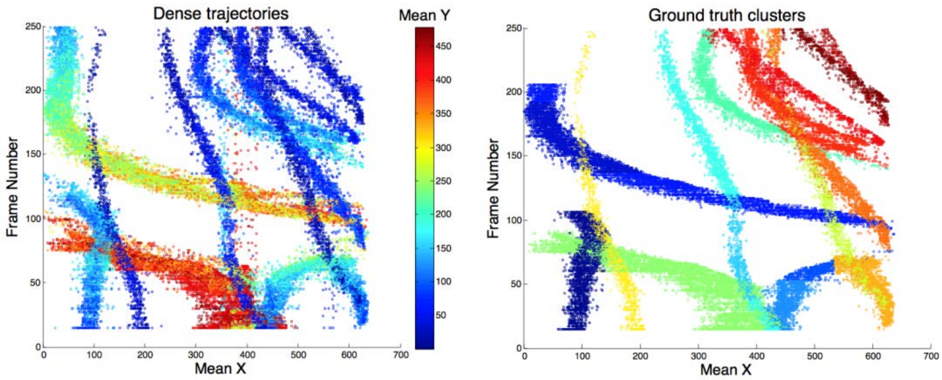


Figure 2. Dense trajectories for a soccer clip. The x-axis is the mean-x position for a trajectory, and the y-axis gives the frame number. **Left**: raw trajectories where colors denote the mean-y position. **Right**: Colors indicate a single player ground truth.

For automatic clustering we use normalized graph cuts [7], as this algorithm naturally allows for non-spherical clusters. The normalized graph cut algorithm [7] is a connectivity-based clustering technique based on a similarity graph. To construct the similarity graph we define the similarity between two trajectories as inversely proportional to the Euclidean distance between their feature vectors: (mean_x, mean_y, frame_nr, dx, dy), where dx and dy denote spatial displacement. In Figure 3 we illustrate results of the clustering on a video from the test set.

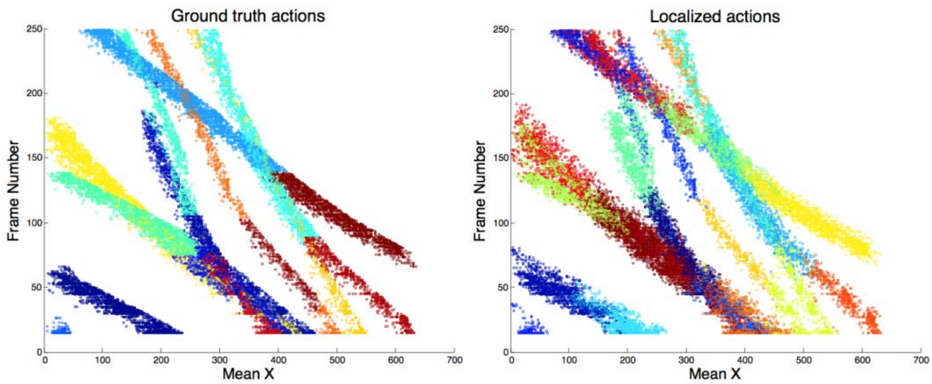


Figure 3. Localization results on a test-video where colors denote clusters. Left: Ground truth. Right: Predicted.

The normalized graph cut algorithm returns a clustering of trajectories. The final output of our action recognition pipeline should be a set of tubelets (Bounding-boxes over time). To obtain tubelets from clusters we simply create a bounding box for the minimum and maximum x and y per frame number. With these clustered tubelets we automatically label each of these clusters using a Support Vector Machine (SVM) classifier with a Histogram Intersection Kernel (HIK) on a Bag-of-Features (BoF) representation, see [1,3,9]. The BoF approach aggregates information from multiple trajectories into a single histogram of prototypes by allowing each trajectory to add a vote to a prototype. The prototypes are obtained by k-means as standardly done. The features here are standard action classification features [9]: Histograms of Oriented Gradietns (HOG), Histograms of Oriented optical Flow (HOF) and Motion Boundary Histograms (MBH), see [9] for details.

Experiments

Data description The video material is shot in full HD with three fixed cameras that together cover the whole soccer field. We recorded two amateur soccer matches at different locations. One match is used to train our models and the other match we keep for testing purpose only. We spatially crop the videos by taking 640x480 pixels around the middle circle of the field. In Figure 4 we show an example of the pitch. Because many actions repeat temporally [8], we split the soccer video into video clips of 10 seconds (250 frames). For the training set we selected 20 and for the test set 12 videos clips, each contain much activity.

We used Mind's Eye Annotation Software (v28) developed by DARPA to annotate the videos with ground truth tubelets. We identified 3 action labels that capture the majority of what players are doing on the field, those actions are walking, running and kicking (the ball). On average a video contains 20 individual actions (tubelets).

Figure 4. Example of the fixed camera on the amateur soccer pitch.

Results

In Figure 5 (left) we show the classification with the confusion matrix showing promising results (79%) between the classes. In Figure 5 (right) we illustrate the statistics that can be obtained as the ration between walking and running. In Figure 6 we show a different type of statistics indicating the location of the performed actions as predicted.

Conclusions

We extract dense trajectories from video clips and cluster the trajectories using normalized graph cuts. The clusters are represented by bag-of-features histograms over the descriptors of trajectories. An SVM classifier predicts action labels for the histograms with an average accuracy of 79% on the test set. Our work is a first step towards fully automatic statistics extraction in amateur soccer.

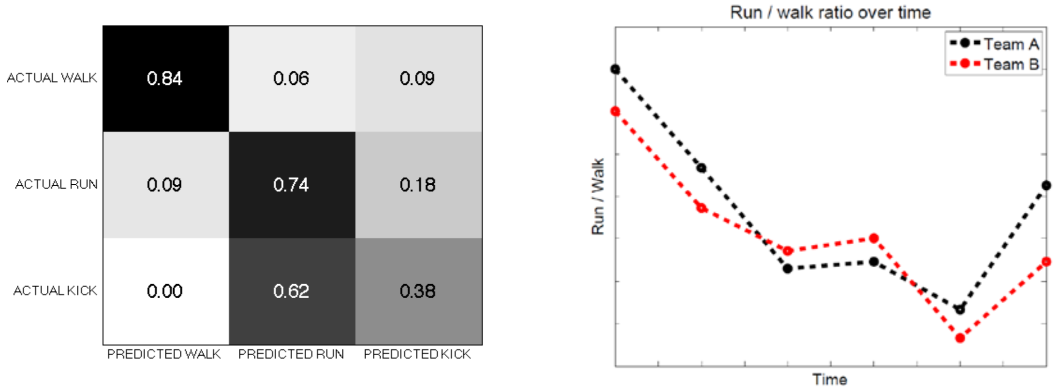


Figure 5. **Left**: Confusion Matrix, avg classification rate: 79%. **Right**: run/walk ratio statistics for the two teams.

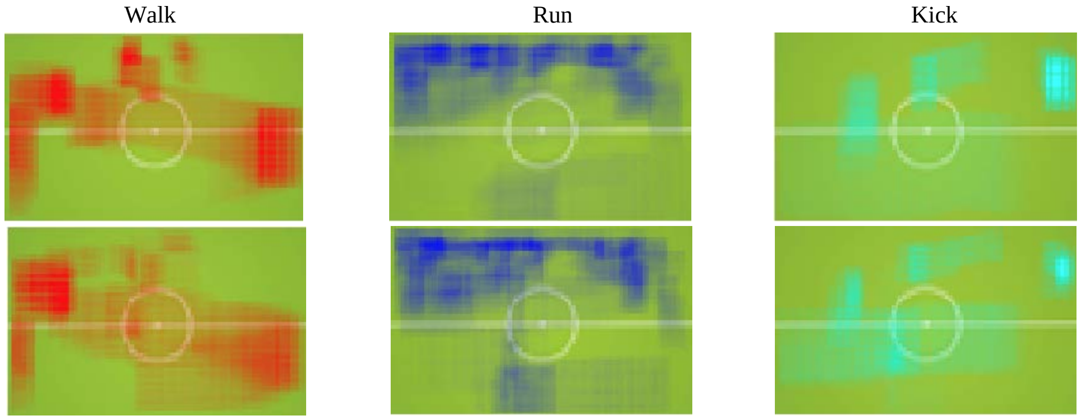


Figure 6. Heatmap of action locations. **Top**: groundtruth actions. **Bottom**: predicted actions.

References

[1] Dalal, N., Triggs, B., Schmid, C., (2006). Human detection using oriented histograms of flow and appearance. *European Conference on Computer Vision (ECCV)*.

[2] Derpanis, K., Sizintsev, M., Cannons, K., Wildes, R., (2010). Efficient action spotting based on a spacetime oriented structure representation. Computer Vision and Pattern Recognition (CVPR).

[3] Everts, I., van Gemert, J.C., Gevers, T. (2013). Evaluation of Color STIPs for Human Action Recognition. *Computer Vision and Pattern Recognition (CVPR)*.

[4] Hoai, M., Lan, Z., de la Torre, F., (2011). Joint segmentation and classification of human actions in video. *Computer Vision and Pattern Recognition (CVPR)*.

[5] Jain, M., van Gemert, J.C., Bouthemy, P., Jegou, H., Snoek, C., (2014). Action Localization by Tubelets from Motion. *Computer Vision and Pattern Recognition (CVPR)*.

[6] Sadanand, S., Corso, J., (2012). Action bank: A high-level representation of activity in video. Computer Vision and Pattern Recognition (CVPR).

[7] Shi, J., Malik, J., (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ***22****(8)*, 888–905.

[8] Van Gemert, J.C., Veenman, C. J., Geusebroek, J. M., (2009). Episode-constrained cross-validation in video concept retrieval. *IEEE Transactions on Multimedia*, ***11****(4)*, 780-785.

[9] Wang, H., Klaser, A., Schmid, C., Liu, C.-L. (2013). Dense trajectories and motion boundary descriptors for action recognition. *International Journal of Computer Vision*, ***103****(1)*, 60–79.

Real-time classification of gorilla video segments in affective categories using crowd-sourced annotations

J. Schavemaker, E. Thomas, A. Havekes

TNO Media & Network Services, Technical Sciences, Delft, The Netherlands

Abstract

In this contribution we present a method to classify segments of gorilla videos[1] in different affective categories. The classification method is trained by crowd sourcing affective annotation. The trained classification then uses video features (computed from the video segments) to classify a new video segment into one of different affective categories: exciting, boring, scary, and moving. As video features we propose to use features based on optical flow. As classification method we propose to use a k-NN classifier for quick relearning possibilities. We validate our method with an experiment with multiple recordings of gorillas from different video cameras and an annotation crowd from within our company.

Key words: affective computing, video annotation, crowd sourcing, classification, optical flow.

Introduction

Live streaming videos are an important part of multimedia that company websites can offer. When broadcasting 'raw' live video the chances are high that the content does not meet the expectations of the viewing audience. This is mainly because the video is not filtered or produced: the video offers no content (e.g. nothing is happening) or the content is in such a format that it is hard to recognize (e.g. overview instead of a close-up of the interesting action). For example, for a zoo that has a digital camera infrastructure that captures the activities of their inhabitants with video cameras and wants to broadcast that live, it is essential that the videos show the activities that will evoke emotions in the audience, preferably content in close-up because that is attractive to the general viewing public.

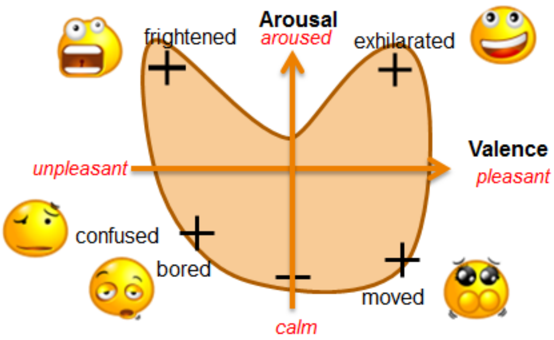
A real-time automatic way of classifying videos into affective categories denoting its relevance would be an attractive asset. This contribution describes a method that learns the classification from labeled examples from crowd sourcing. The classification learns to link measurable image features with affective viewing categories.

Related Work

Image annotation is the process by which a computer system or person assigns metadata in the form of captioning or keywords to a digital image. Affective annotation records the viewer's emotion; in general it is a subjective annotation. A common model of describing emotions is the two-dimensional arousal-valence diagram of Russel [7], see Figure 1. Common approaches along this model are to define one emotion for each quadrant of the arousal-valence diagram or to use negative, neutral and positive valence values [1].

Hanjalic et al [4] present an affective video content representation and modeling. Their approach is content-based: they compute affective annotations from the video itself. In their representation viewer emotions are described in a 2D arousal space. They note that motion in video increases arousal but valence is not modified. Motion, audio and shot lengths are used as low-level features in this paper. Hu [5] gives a broader overview of content-based features, see Figure 1. Soleymani et al [3] describe the development of a viewer-reported boredom corpus by crowd sourcing affective annotation of video. Using Mturk, crowdsourcing provides an effective means of collecting the viewer affective response annotations needed to create a corpus to be used in the development of automatic prediction of viewer reported boredom.

[1] We would like to thank Apenheul for use of their video material and additional support.



Domain	Features	
Video Structure	Shot boundary detection	Threshold-Based Approach:
		Statistical Learning-Based Approach:
	Key frame extraction (frames that best reflect the shot contents)	Sequential Comparison Between Frames
		Global Comparison Between Frames
		Reference Frame
		Clustering
		Curve Simplification
		Objects/Events
	Scene Segmentation	Key Frame-Based Approach
		Audio and Vision Integration-Based Approach
		Background-Based Approach
Picture	Color	color histograms, color moments, color correlograms, a mixture of Gaussian models, etc
	Texture	Tamura features, simultaneous autoregressive models, orientation features, wavelet transformation-based texture features, co-occurrence matrices, etc.
	Shape	detect edges then describe the distribution of the edges using a histogram
Object	dominant color, texture, size, etc., corresponding to the objects	
Motion	Statistics	model the distribution of global or local motions of points in the video
	Trajectory	extracted by modeling the motion trajectories of objects in videos
	Objects' Relationship	relationships between objects using a symbolic representation scheme

Figure 1. (left) Arousal-valence diagram. (right) Different types of content-based features by Hu [5].

Method description

In this section we describe our classification method. The method is part of the following scheme:

phase 1 : affective annotation of video segments by crowd sourcing;
phase 2 : learning video classification from annotations and features;
phase 3 : real-time classification of new video segments.

Annotation of video segments by crowd sourcing

The first phase consists of collecting videos and annotating videos by crowd sourcing. For the experiments we have collected videos of the Gorilla Island of the Apenheul[2]. The Gorilla Island is recorded with eleven high-definition cameras capturing different views of the Gorilla Island; see Figure 2 for some example views. Our video collection consists of videos that have been captured for a number of days. The videos have been segmented and transcoded (using FFmpeg[3]) into fragments of 10 seconds, from which a selection is made.

Figure 2. Different camera views of the Gorilla Island in the Apenheul (images courtesy Apenheul).

The annotation of the videos is performed with the help of an annotation crowd consisting of more than 100 participants within our research company and a web-based annotation tool. Every participant views a number of randomly selected video segments from the video pool. The participant selects the appropriate affective category that best fits her/his emotional state after viewing the video segment, see Figure 4. This annotation process is implemented as a web-based video player with additional functionality to record the annotation. Every participant is also asked to enter age, gender and nationality. This allows future analysis for tailoring classification to different viewing groups.

Learning video classification from annotations and features

[2] http://www.apenheul.nl/ [Last accessed on January 27, 2015]
[3] http://www.ffmpeg.org/ [Last accessed on January 27, 2015]

To train a classifier that classifies a video segment into an affective category the relationship between annotation and video content must be made. In order to do that we construct a content-based feature vector for every video segment that captures the video content. In our method we have chosen for optical-flow features: the hypothesis is that optical flow correlates with arousal (amount of motion) and valence (type of motion) of the viewer. Optical flow is commonly used to classify (human) activities using a histogram approach like histograms of oriented optical flow (HOOF) [2]. This approach bears similarities with the popular histogram of oriented gradients (HOG) approach by Dalal et al.

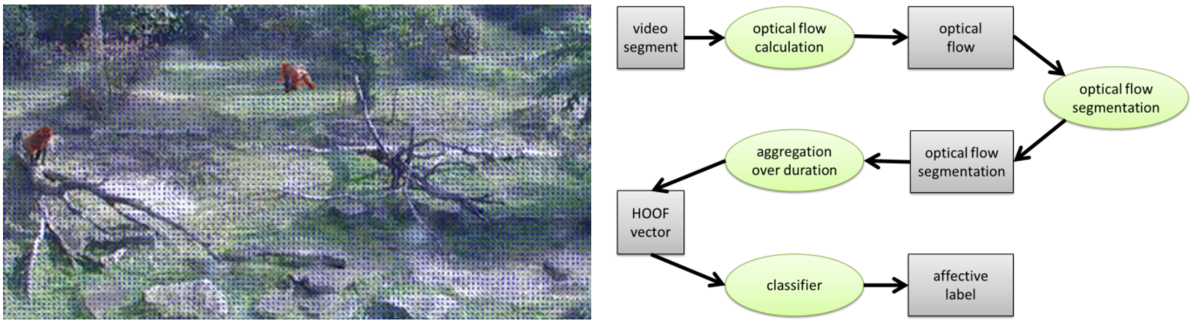


Figure 3. (left) Optical flow segmentation, blue: background, red: foreground. (right) Real-time classification diagram.

For our experiments optical flow is computed for every video frame using the Dual TV L1 optical-flow algorithm described in [7] and as implemented in the OpenCV[4] library. After calculation, the optical flow is segmented to extract the foreground optical flow actions and not let noise and clutter distort the histogram. To do that, we compute per camera the mean and standard deviation of optical flow magnitude per image block for a set of videos without gorillas. When an optical flow vector is larger in magnitude than corresponding mean magnitude plus 3 times standard deviation the optical-flow vector contributes to the feature computation. The segmented optical flow is aggregated over the duration of the complete video segment to create a histogram of oriented optical flow (HOOF). The HOOF vectors with the associated labels are used to train the classifier.

Real-time classification of new video segments

Real-time classification of new video segments (without annotation) proceeds much in the same way as learning. For a new video segment, optical flow is computed, segmented and aggregated into a HOOF feature vector. The HOOF feature vector is fed into the classifier to obtain a classification of the video in one of the categories.

Experiments

For the experiments we use a dataset with video segments of 3 cameras. The total number of video segments for every camera equals: 505, 461, and 167. The crowd annotation consists of 1623, 644, and 532 ratings (per video segment multiple ratings exist, ranging from 1 to 10 raters). The inter-rater reliability has been computed from the segments having 3 ratings. The Fleiss' kappa [6] equals 0.26 in this case, corresponding with a 'fair agreement' according to Landis and Koch[9]. The dominant rating is 'boring', followed by 'exciting', 'moving' and 'scary'. Ratings differ per camera, for example camera 1 (the feeding area) has about 65% boring, 24% exciting, 11% moving and less than 1% scary. Videos for which there is no majority vote of the rates are removed from the set because of the fair rater agreement. Next, the dataset is divided in a train and test set of video segments for each camera and per camera a classifier is trained. The train set per camera consists of the video segments (and computed feature vectors) with the corresponding labels from the raters. So, one video segment in the train set can occur multiple times with all the labels from the different raters. For the test set, every video segment has one ground truth label that equals the majority vote of its ratings. As a classifier we use a k-NN classifier with a major voting scheme on the nearest neighbors of the sample (k = 11). Figure 4 presents

[4] http://opencv.org/ [Last accessed on January 27, 2015]

the classification accuracies for the different cameras and the confusion matrix (actual versus predicted classification percentages) for camera 1.

camera	1	3	7
accuracy	84,80%	94,04%	87,90%
#train videos	171	218	58
#test videos	171	219	58

	moving	boring	exciting
moving	1%	3%	1%
boring	0%	77%	5%
exciting	0%	6%	8%

Figure 4. (left) Web-based annotation tool. (right top) Classification accuracies. (right bottom) Confusion matrix camera 1.

Figure 4 shows high accuracies but that is mainly because the dominant class is 'boring'. From the confusion matrix we may deduce that there is some accuracy for 'moving' and 'exciting' albeit marginal. This is largely because 'moving' and 'boring' will overlap in features space as well as to some extent 'exciting' and 'boring'.

Conclusions

In this contribution we have presented a method to classify segments of gorilla videos in different affective categories. The classification method uses crowd sourcing annotation which is combined with features computed from the video segments to classify a new video segment into different categories: exciting, boring, scary, and moving. We validated our method with a small experiment with gorilla videos and a small crowd. The computed optical-flow features show promising classification results. Future work will focus on a larger experiment.

References

[1] van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unraveling the mystery of emotions.* CTIT Ph.D.-thesis series No. 11-204, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.

[2] Chaudhry, G.H.R., Ravichandran, A., Vidal, R. (2009). Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions. *CVPR*.

[3] Soleymani, M., Larson, M. (2010). Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. In *Proceedings of the Workshop on Crowdsourcing for Search Evaluation (SIGIR'10)*. July 19–23, Geneva, Switzerland.

[4] Hanjalic, A., Xu, L.-Q. (2005). Affective video content representation and modelling. *IEEE Transactions on Multimedia*, ***7****(1)*, 143–154.

[5] Hu, W., Xie, N., Li, L., Zeng, X., Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, ***41****(6)*, 791–819.

[6] Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, ***76****(5)*, 378–382

[7] Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, ***39****(6)*, 1161–1178.

[8] Zach, C., Pock, T., Bischof, H. (2007). A duality based approach for realtime TV-L1 optical flow. In Proceedings of the 29th DAGM conference on Pattern recognition, p. 214-223. Berlin/Heidelberg, Germany: Springer-Verlag.

[9] Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, ***33****(1)*, 159–174.

Social media offers new insights into human and animal behaviour: How to harness them scientifically

C.C. Burn

Centre for Animal Welfare, The Royal Veterinary College, North Mymms, UK
cburn@rvc.ac.uk

Background

New, inexpensive opportunities for studying the behaviour of humans, animals and almost anything else, from machines to geophysical phenomena, are arising from social media sources [1]. The most obvious sources of behavioural data are video repositories, YouTube being by far the largest to date, and thus they are the focus of this work. However, repositories of photographic or written descriptions of behaviour can also be informative, with written comments having special relevance to studies of human attitudes and trends.

Subjects so far have ranged from the fundamental to the applied in both human and animal behaviour, covering fields including medicine, psychology, national defence, animal behaviour and veterinary medicine. Seemingly, the first studies to code behaviour from YouTube were published in 2009 [2, 3], but only a small handful of studies have done so each year since.

This paper presents an encouraging but cautionary message. Social media can provide vast amounts of data ripe for surveying (e.g. about 100h of video footage is uploaded to YouTube every minute), but without careful design of such surveys, conclusions can be prone to bias and misinterpretation. Yet, prejudice against these data sources, based on a belief that they must be somehow 'unscientific', unnecessarily dismisses this rich new information. Here I discuss the strengths and limitations of surveying behaviour from online video repositories, illustrating the potential of the methodology with examples.

Strengths of online video data

While some researchers are initially sceptical about using data from social media, there are many parallels between its strengths and limitations and those of questionnaire data and epidemiological surveys, as compared to direct experimental manipulations. Some shared strengths are as follows:

- Large sample sizes may be possible. For example, the largest direct observation sample size of dogs that chased their tails was 32 [4], compared with 446 in a questionnaire [5], and >3500 hits in a YouTube survey, of which 400 were sampled [6].

- The method is inexpensive. Videos, e.g. on YouTube, Vimeo and DailyMotion, are currently free to view.

- The method is amenable to exploring human attitudes. On YouTube, attitude data can be collected from the uploader description of the video, viewer comments, and via observing the humans in the videos directly. Examples include: a study 115 videos of parents interacting with toddlers while playing with toy kitchens, revealing implicit gender- and dietary-biases in the subjects' conversations, actions and choices of toys used [7]; and a study showing that 73 videos portraying negative views of immunisation received more viewings and higher star ratings than 49 videos portraying positive views [8].

- Rare and/or fleeting events can be studied. On video sharing sites, this relies on the phenomenon of interest being amenable to video capture. It also requires that it has searchable and widely used names or contexts, because without this, relevant videos will not be easily located.

  Tail chasing is a good example, because most dogs perform it rarely and briefly, so it is difficult to research by direct observation (except in dogs who perform it compulsively) [6]. Yet the behaviour is well known, so when people see it, they may well decide to video and upload it for others' amusement. 'Tail chasing' is also searchable, unlike another potentially compulsive canine behaviour, 'fly catching'/ 'fly snapping'/ 'fly biting': technical terms that the general public would not necessarily use.

Data collected from online videos has some further strengths that questionnaires usually do not. For example:

- The videos allow researchers to quantitatively record behaviour directly, in just the same ways as they would any other videos. Researchers can apply their own objective criteria and descriptions of the relevant behaviour, and can measure frequencies, durations etc. directly. They can see for themselves what viewers are commenting on. An example is the objective assessment of true 'motor entrainment' to music in animals described as 'dancing' by uploaders; that study reported motor entrainment to occur only species capable of vocal mimicry (eight parrot species and one elephant species), but never in the >600 video sample of non-mimicking species [2].

- Unlike questionnaires, the data are not prompted by the study itself. This points to a paradoxical situation whereby data are submitted to public viewing and yet the process of data collection is unobtrusive. This can remove bias that is created by the phrasing of questionnaires. For example, in the tail chasing study, >20% of videos showed people encouraging the behaviour in a manner that could compromise dog welfare (e.g. pulling the tail, attaching objects to it, or 'growling' at the dog). Also uploaders were over six times more likely to describe dogs as 'stupid' if they tail-chased compulsively than if they performed it rarely. Another example is the study by Lynch [7] showing parents reinforcing gender stereotypes and setting unhealthy examples of food choice whilst playing with their toddlers. In both examples, a questionnaire would likely have prompted people to reflect on their actions and how they might be perceived, whereas – perhaps surprisingly – deciding to publicly display video footage evidently seems not to prompt reflection in the same way.

- YouTube with its vast viewing figures has a special quality in that it not only offers insight into the real world, but it can also influence it. Many viewers use YouTube as a source of information, so the behaviour they see and the comments they hear or read, may alter their perceptions, reactions to, or participation in that type of behaviour if they encounter it in the real world. Linkletter et al. [9] found that 65 videos of an asphyxiation 'game' in teenagers were collectively viewed 173550 times, and warned of the potential of these videos to normalize and even propagate the behaviour. A similar phenomenon was seen with self-harm videos, where the top 100 videos were viewed >2 million times, and selected as user 'favorites' over 12 000 times [10]. A further sinister example is the use of YouTube for promoting violent terrorism [3]. Positive influences are also possible via videos that accurately raise awareness of important issues or teach beneficial skills to viewers.

Limitations of online video data and what to do about them

Questionnaires and epidemiological methods have been used for long enough now that there is an understanding of how best to conduct them and to interpret the resulting data [e.g. 11]. Many of the same principles can actually be applied to social media data as well. Again many of the limitations of social media data are shared with questionnaires and many other surveys, namely:

- Non-random sampling. Like most respondents to questionnaires, social media data are only collected from people who voluntarily submit their 'data'. The normal way to deal with this in questionnaires is to describe the demographics as far as possible, and to discuss any likely factors that might have biased the sample population. This approach can be used for social media also. On YouTube, clicking on the uploader username reveals volunteered information on their age, sex and nationality. Conclusions should be restricted to populations resembling the sampled one and any likely biases discussed. Depending on the question being asked in the study, the biases might restrict the conclusions very little, and in some cases an appropriate control group may be possible to help correct for it, e.g. the prevalence of 'dancing' in the vocal mimicking species compared with the non-mimics [2].

- It is good practice to assess any data for internal validity, to ensure that what is being measured actually reflects the intended construct. The validity of noisy data like that often collected from social media may be questioned, so it is wise to take several measures relevant to the phenomenon of interest if possible. Statistical tests can then be used to assess whether the measures correlate with each other in the manner expected. For example, two measures of compulsive behaviour in the tail-chasing dogs correlated positively, while they correlated negatively with playful behaviour [6].

- Untruthful descriptions or opinions are always a risk with questionnaires, and this may be no less true with social media studies. Social media studies have the aforementioned advantage of being unobtrusive, but when people upload videos and comments, they intend them to be viewed by 'the public' (they proactively select this option from a drop down list on YouTube). Therefore, an element of self-consciousness is still intrinsic to the data. Case by case judgement is required as to how much the phenomena under study will be influenced by this, and it should be discussed with respect to interpreting the data.

  As with questionnaires and some direct observations, the background context is often unknown. In the case of videos, many circumstances surrounding the recording are unknown, and the ground truth for the behaviour recorded can be hard to determine (e.g. a video starts at a certain moment and it remains unknown what happened before that moment).

  Falsified footage is a new concern specific to audio-visual data, but since most videos uploaded to YouTube and other video-sharing websites are amateur, alterations are currently usually quite obvious and easy to identify. Again, certain research questions will be more prone to this problem than others.

- While some data collected from social media will be standardized and structured across the sample population (e.g. each video has a title, category and number of views, and users have demographic information), much of it will not be. There may be many missing data for some variables. Videos may be of different lengths and different qualities, and text will normally be free form. Setting inclusion criteria helps quality control which videos or other data are included in the sample, as does recording data in a form that enables extraneous variation to be taken into account in analyses (e.g. recording behaviour frequencies relative to the video duration, or describing vocalization data only for the subset of videos that had audio associated with them). With regards to free text, there are numerous techniques from the field of social science that can be applied to help quantify such qualitative data, such as thematic analysis [e.g. 12]. Researcher bias can affect qualitative data, because interpretation occurs both before and after any statistical analysis, but again methods to control this, and to make any likely influences transparent, can be employed [e.g. 13].

Discussion

In conclusion, using data from social media offers some exciting and sometimes quite profound new opportunities for studying behaviour, as long as it is collected and interpreted with due scientific caution and integrity. In particular, shared video repositories can enable large sample sizes of sporadic behaviour episodes to be measured quite directly, which would be difficult or impossible to do any other way, and can reveal human attitudes towards those behaviours.

References

[1] Rault, J.-L., Elmore, M.R.P., Biehl, D.J., Russell, M.A., Garner, J.P. (2013). The world is a natural laboratory, and social media is the new petri dish. *Ethology*, ***119***, 803-806.

[2] Schachner, A., Brady, T.F., Pepperberg, I.M., Hauser, M.D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, ***19***, 831-836.

[3] Weisburd, A.A. (2009). Comparison of visual motifs in Jihadi and Cholo videos on YouTube. *Studies in Conflict and Terrorism*, ***32***, 1066-1074.

[4] Blackshaw, J.K., Sutton, R.H., Boyhan, M.A. (1994). Tail chasing or circling behavior in dogs. *Canine Practice,* ***19***, 7-11.

[5] Goto, A., Arata, S., Kiyokawa, Y., Takeuchi, Y., Mori, Y. (2012). Risk factors for canine tail chasing behaviour in Japan. *The Veterinary Journal,* ***192***, 445-448.

[6] Burn, C.C. (2011). A vicious cycle: a cross-sectional study of canine tail-chasing and human responses to it, using a free video-sharing website. *PLoS ONE,* ***6***, e26553.

[7] Lynch, M. (2010). Playing with food. A novel approach to understanding nutritional behaviour development. *Appetite,* ***54***, 591-594.

[8] Keelan, J., Pavri-Garcia, V., Tomlinson, G., Wilson, K. (2007). YouTube as a source of information on immunization: a content analysis. *Journal of the American Medical Association,* ***298***, 2482-2484.

[9] Linkletter, M., Gordon, K., Dooley, J. (2010). The choking game and YouTube: a dangerous combination. *Clinical Pediatrics,* ***49***, 274-279.

[10] Lewis, S.P., Heath, N.L., St Denis, J. M., Noble, R. (2011). The scope of nonsuicidal self-injury on YouTube. *Pediatrics,* ***127***, e552-e557.

[11] de Vaus, D. (2001). *Surveys in Social Research* (5th ed.). Routledge.

[12] Braun, V., Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* ***3***, 77-101.

[13] Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet,* ***358***, 483-488.

Automated classification of rat social behavior

S.M. Peters [1,2], I.J. Pinter[2], R.C.de Heer[2], J.E. van der Harst[2], B.M. Spruijt[1]

[1]Department of Behavioral Biology, Faculty of Science, Utrecht University, the Netherlands
s.m.peters@uu.nl
[2]Delta Phenomics B.V., Schaijk, the Netherlands

Introduction

Several psychopathologies affecting social behavior such as autism disorders, major depression disorder and schizophrenia are detrimental diseases for which currently no cure exists and much of the pathology is still unknown. The development of valid animal models and reliable test paradigms are crucial elements in the process of understanding the underlying mechanisms of these disorders. Rodents are, similar to humans, very social animals. This aspect, however, has been neglected in behavioral neuroscience and drug discovery. While the quantification of rodent social behavior has long relied on manual observations, recent advanced software and/or hardware technologies e.g. [1-5] make it possible to automatically monitor multiple animals and their social interactions. This requires a different approach in analyzing the data that is generated by these systems. Here, we propose a method to analyze social interactions of rat pairs. Since this method is not relying on subjective human interpretation, but on computerized data, it is more objective and less labor intensive. Also, we will argue here that our method is very valid in ethological terms, because criteria differentiating various behaviors are based on the occurrence (using individual frequency distributions) of the animal's own behavior.

Methods

Pairs of male rats of 5 weeks old were placed in a relatively large PhenoTyper® cage of 90 cm x 90 cm (Noldus Information Technology, the Netherlands) and allowed to freely interact (i.e. social interaction test). At this age rats are known to display the highest levels of social (play) behavior (e.g. [6, 7]). All testing was done during the dark phase (that is the active phase) of the animals under red light conditions and all animals were habituated to human handling. Before each social interaction test, animals were marked red or black (permanent marker, Edding, Germany) for the software to recognize both individuals based on their marking. The black marking is visible in the video, while the red marking is not visible (due to the infrared light conditions, see below). In this way both animals experience the same practical procedure, but the software recognizes a marked and an unmarked animal; a similar procedure is used by [8, 9]. Incorrect identity swops (in the initial version of the software this was about 50% of the samples) made by the software were manually corrected after video tracking. The experiments were performed in adherence to the legal requirements of Dutch legislation on laboratory animals (Wod/Dutch 'Experiments on Animals Act') and were reviewed and approved by an Animal Ethics Committee ('Lely-DEC').

From each social interaction test session, top view video recordings were made with a top-unit placed on top of the PhenoTyper, containing an infrared sensitive camera (CCD 1/3" SONY SUPER HAD CCD black/white) and IR-filter (type Kodak 87C). In the development of our method we were inspired by Golani and coworkers that have for long been working on automatic classification of rodent behavior, e.g. [10]. Their approach involves the Gaussian expectation maximization method to search for natural modes in the data based on frequency distributions (see [11] for an extensive explanation of this method). After tracking the animal from video recordings with automatic tracking software (EthoVision XT 8.0, Noldus Information Technology, the Netherlands), raw data containing x- and y-coordinates was exported and further analyzed in MatLab® R2012b (The MathWorks, United States). First, raw track data was smoothed using a robust LOWESS filter with a 1-s time window in order to remove noise. Then, parameters such as the distance moved, speed of movement and distance between two animals were determined. The tracking data was divided between movement bouts and stops (i.e. arrests). Therefore, data was filtered with a repeated running median using a one dimensional median filter with order 7, 5, 3 and 3 respectively. Based on visual inspection of the tracks of the animals and video images the threshold for a stop was set at 0.07 cm between 2 samples lasting at least 0.16 seconds. Next, a density estimator (i.e. smoothed frequency distribution) of the maximal velocity of each movement bout was made. On these density estimator plots the best Gaussian curves to represent the empirical data were fitted using an expectation maximization (EM) method. Subsequently, the intersection of these curves determined the different “modes” or velocity categories in which the animals moved. Additionally, the same analysis was done on the distance between the animals, with the difference that this parameter was calculated for each sample in the track data. Also, for this distance between parameter the intersections of the Gaussian curves that best represented the empirical data were used to determine the different “modes” of proximity. Thereafter, we combined both movement and proximity modes together to obtain specific approach and avoidance behaviors characterized by velocity and distance.

Results and Discussion

Our method revealed that in our setup the movement of juvenile rats (when freely socially interacting) can be divided between movements with either low velocities or high velocities (average threshold of 34 cm/s). In addition, we found that distance between rat pairs can be classified in 3 different proximities categories: in -very-close proximity, in proximity and not in proximity (average thresholds: 11 and 31 cm, respectively). Combining these two modes (velocity with proximity) we could clearly identify for example behavioral bouts in which the animals are moving with high velocity and in close proximity, representing chasing behavior. Currently, we are also investigating social interactions of adult rat pairs with this method and we are applying our method on existing models used to examine social deficits, for example phencyclidine induced social impairments.

Using automated social parameters based on coordinates of the animals is not completely new. Sams-Dodd (1995) used a similar method to create the automated parameter ‘social interaction’ [8]. This parameter was made with a fixed threshold; animals were regarded to be in proximity when their center of gravity points are within 20 centimeters from each other. He mentioned that: “selection of this criterion value of 20 cm is based on systematic variation of this parameter from 0 to 50 cm” and “the value of 20 cm resulted in the least variation in the data”. It is exactly this variation that we now use to classify the different categories of proximity. Important benefit of this approach is that the threshold is not created artificially by limiting variation, but based on the animal’s own behavior and natural variation within this behavior. In addition, it is now also possible to combine the information represented by speed of movements of the animals with the categories of proximity.

The next step is to apply this method on data from continuously monitored group-housed rats to study their social behavior displayed in a home-cage environment for more precise registration and thus enhanced differentiation of rodent social behavior. This is also beneficial for increasing the translational value of social behavior when employed as an indicator for assessing treatment efficacy in animal models for psychopathology involving social behavior.

Acknowledgements

The large PhenoTyper cage was developed as part of the SenseWell project. We acknowledge our partners in this consortium (Metris BV, Noldus Information Technology BV, Telemetronics Biomedical BV, BioMetris Wageningen University, Faculty of Veterinary Medicine Utrecht University and EEMCS TU Delft) for their support. This research was partially financed by a grant from Agentschap NL (NeuroBasic-PharmaPhenomics) and ICT Regie / NWO (SenseWell).

References

[1] de Chaumont, F., Dos-Santos Coura, R., Serreau, P., Cressant, A., Chabout, J., Granon, S., Olivo-Marin, J.-C. (2012). Computerized video analysis of social interactions in mice. *Nature Methods, **9**, 410-417.*

[2] Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., Branson, K. (2013). JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, ***10***, 64-67.

[3] Giancardo, L., Sona, D., Huang, H., Sannino, S., Managò, F., Scheggia, D., Papaleo, F., Murino, V. (2013). Automatic Visual Tracking and Social Behaviour Analysis with Multiple Mice. PLoS ONE, ***8****(9)*, e74557.

[4] Weissbrod, A., Shapiro, A., Vasserman, G., Edry, L., Dayan, M., Yitzhaky, A., Hertzberg, L., Feinerman, O., Kimchi, T. (2013). Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nature Communications*, ***4***, 2018.

[5] Shemesh, Y., Sztainberg, Y., Forkosh, O., Shlapobersky, T., Chen, A, Schneidman, E. (2013). High-order social interactions in groups of mice. *eLife*, ***2***, e00759.

[6] Panksepp, J. (1981). The ontogeny of play in rats. *Developmental Psychobiology*, ***14***, 327-332.

[7] Pellis, S.M., Pellis, V.C. (2007). Rough-and-tumble play and the development of the social brain. *Current Directions in Psychological Science*, ***16***, 95-98.

[8] Sams-Dodd, F. (1995). Automation of the social interaction test by a video-tracking system: Behavioural effects of repeated phencyclidine treatment. *Journal of Neuroscience Methods*, ***59***, 157-167.

[9] Spruijt, B.M., Hol, T., Rousseau, J. (1992). Approach, avoidance, and contact behavior of individually recognized animals automatically quantified with an imaging technique. *Physiology and Behavior*, ***51***, 747-752.

[10] Benjamini,Y., Lipkind, D., Horev, G., Fonio, E., Kafkafi, N., Golani, I. (2010). Ten ways to improve the quality of descriptions of whole-animal movement. *Behavioral Brain Research*, ***34***, 1351-1365.

[11] Drai,D., Benjamini,Y. and Golani,I. (2000) Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods*, ***96***, 119-131.

The development of a non-invasive behavioral model of thermal heat stress in laboratory mice (*Mus musculus*)

J.T. Mufford[1], M.J. Paetkau[1], N.J. Flood[1], G. Regev-Shoshani[2], C.C. Miller[2], J.S. Church[1]

[1]Thompson Rivers University, 900 McGill Road, Kamloops, British Columbia V2C 0C8, Canada
[2]Department of Medicine, Respiratory Division, University of British Columbia, Vancouver, British Columbia V5Z 3J5, Canada

Previous research investigating physiological and behavioral responses of heat stressed mice have often relied on invasive biotelemetry methods in order to collect crucial data such as core body temperature [1, 2]. Biotelemetry generally requires surgical implantation of a telemeter, which subjects the animal to surgical distress and discomfort along with the risk of infection [2]. Furthermore, the invasiveness of a surgical procedure may influence the behavior of the animal test subjects.

We wished to develop a method for studying thermal heat stress in mice that did not require invasive surgical procedures and that allows precise control of the ambient heating conditions while measuring the external surface temperature of the animal test subject. The new experimental design was configured such that both the behavior and surface temperature of mice can be continuously monitored automatically while undergoing different heating regimes in a modified 30x30cm open field (Med Associates Inc, St. Albancs, VT) (Figure 1). To induce heat stress in mice, one wall of the open field was modified to include a ceramic heat emitter (Hagen, Baeie d'Urfé, QC Canada) that was directed toward the inside of the field, providing an increase in the ambient air temperature that was held constant in our experiments at 36°C. The ceramic heat emitter was controlled with a variable voltage transformer (R. Mack & Co. Ltd., Vancouver, BC, Canada) and a power monitor (P3 International, New York, USA). A cooling fan (KDE1205PFB2-8, Thermal FX, Gardner, NV, USA) connected to a DC Power Supply (HP, Palo Alto, CA, USA) was also installed in the bottom corner of the modified wall of the open field for air circulation. A custom-built cage top was fabricated, to keep the heated air trapped in the open field box, while still allowing the use of the infrared thermography (IRT) camera (FLIR E60 infrared camera, FLIR Systems Inc., Wilsonville, USA) external to the open field. The key component of the cage top included a sheet of thin plastic stretched tightly on a wooden frame. This allowed us to simultaneously record the behavioral activity of the mice with a colour CCT camera (Panasonic WV-CP504, Newark, USA) and the surface temperature of both the mouse while in the open field, and the surface temperature of the open field with the IRT camera. The surface temperatures were recorded by streaming radiometric MPEG to a computer using FLIR Tools+ software (FLIR Systems Inc., Wilsonville, USA) (Figure 2).

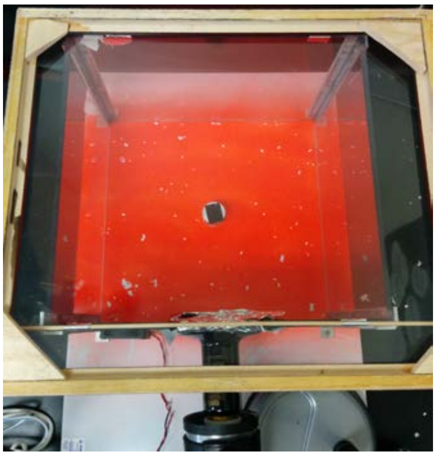


Figure 1. Open field with a custom built plastic top and a modified wall.

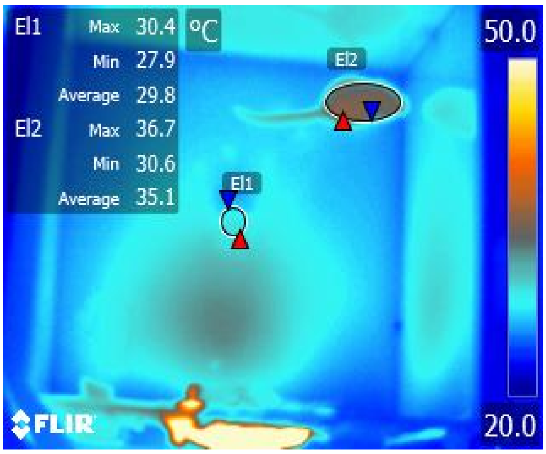


Figure 2. Snapshot of streamed IRT video recording. Surface temperature of the mouse and the open field are simultaneously being measured.

A thermocouple (LM 35 Heat Sensor, Texas Instruments, Dallas, Texas, USA) was installed at the bottom of the cage, raised 3cm from the bottom and enclosed with aluminum, for secondary temperature validation. The surface temperatures of the open field was also recorded by taking the IRT measurement of a piece of vinyl electric tape with known emissivity (Scotch Super 88, 3M, St. Paul, Mn, USA) attached to the thermocouple. The temperature recorded by both the thermocouple as well as the IRT measurement of the tape were nearly identical. The novel plastic cover also enabled other environmental variables to be investigated such as the effect of electromagnetic radiation of visible light on different mouse coat colors by using LED video lights (Gentec International, Markham Canada) placed adjacent to our open field.

The colour CCT camera recorded mice activity through Media Recorder 2.5 (Noldus, Wagenigen, The Netherlands) using a Euresys Picolo U4 H.264 capture card (Euresys, San Juan Capistrano, USA). The behavior of the mice in response to thermal loading was automatically tracked using Ethovision 10.0 software (Noldus, Wagenigen, The Netherlands) (Figure 3) to determine the total distance travelled and velocity of the mice inside the open field using the CCT camera video file. In general, the total distance travelled by mice in the heated open field was far less than that of the mice in the control open field, which was maintained at room temperature in the laboratory (22°C). Ethovision was also used to generate heat maps representing the areas in the open field where the mice had spent the most time (Figure 4). Four continuous data streams (ambient air temperature of the open field measured by a thermocouple; surface temperature of the open field measured by IRT mpeg radiometric video of the electrical tape, surface temperature of mice measured by IRT mpeg radiometric video, along with the behaviour of mice captured by avi video by the CCT camera) can now be readily statistically compared. The development of this novel non-invasive model of heat stress improves the animal welfare of the study animals, while still allowing for very detailed control of both heat and lighting factors, along with unimpeded behavioral responses of the animals.

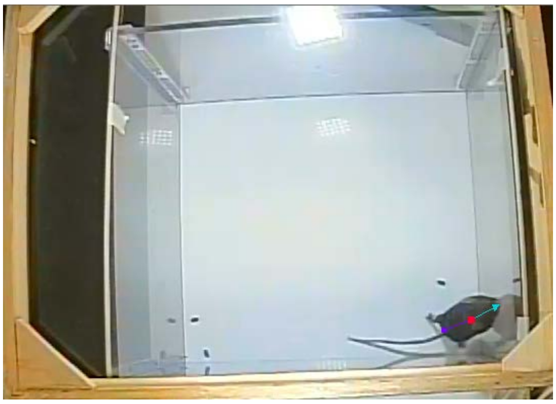


Figure 3. Automated mouse tracking in Ethovision 10.0. Mouse center, nose and tail points are tracked to determine velocity and total distance travelled.

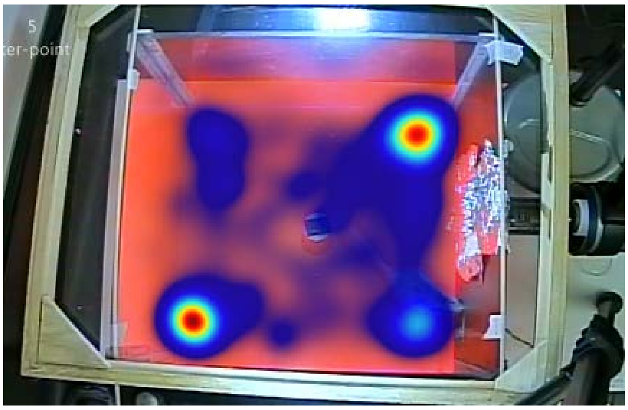


Figure 4. Heatmap representation of the areas in the open field where the mouse had spent the most time. Duration is scaled from blue, representing the least duration, to red, representing the highest duration.

Ethical Statement

This study was approved by the Thompson Rivers University Animal Care Committee and followed the Canadian Council on Animal Care Guidelines.

References

[1] Leon L.R., DuBose D.A., Mason C.W. (2005). Heat stress induces a biphasic thermoregulatory response in mice. *American Journal of Physiology*, ***288***, 197-204.

[2] Gaskill B.N., Gordon C.J., Pajor E.A., Jeffrey R.L., Davis K.D., Garner J.P. (2013). Impact of nesting material on mouse body temperature and physiology. *Physiology & Behavior*, ***110-111***, 87-95.

Measuring Facial Expression and Emotional Experience under Diverse Social Context in a Negative Emotional Setting

Ying Jiang

Dep. Interactive Media - Human Factors, Fraunhofer Heinrich Hertz Institute, Berlin, Germany. ying.jiang@hhi.fraunhofer.de

Abstract

This paper presents an empirical study of the social context effect on facial expressions in a negative emotional setting. In this study participants watched sadness-inducing film clips under different social context (alone, with stranger or with friend) and their facial expressions were coded with the Observer XT using the Facial Expression Coding System (FACES; [1]). According to Fridlund's Behavioral Ecology View [2], the presence of an intimate social interactant may facilitate negative expression. Yet the assessment and evaluation of facial behavior and self-reports of emotion revealed that negative expression did not vary as a function of sociality whereas positive expression was facilitated with the presence of a friend despite the negative valence of the stimuli. Furthermore participants reported being less sad in the presence of a friend. The finding may imply that the effect of the social context may show a different pattern regarding negative and positive facial expression in different emotional settings.

Introduction

Psychologists hold two different views about human facial expression. The belief that certain facial behaviors express or signal emotion has been a predominant account for the past 50 years [3] [4] [5]. It is assumed that there are a number of basic emotions, which are innate and universally the same independent of cultural differences. For each basic emotion, there is a prototypical pattern of facial expression. Once the prototypical facial expression appears, it can be assumed that a certain emotion is evoked. However, some psychologists doubted the existence of an emotion-expression link. Most representative of this is Fridlund's Behavioral Ecology View [2]. He argued that facial expressions are solely tools for communicating social motives to specific addressees, not direct manifestations of the inner emotional state. Empirical evidence for the latter view arises mostly from the observation and measurement of facial movement under the experimental manipulation of social context [6] [7] [8]. Most of the research had been focused on *positive* emotions, reaching the conclusion that people tend to show more positive expressions with the increase of sociality, whereas their emotional experience does not covary with the change of social context. In addition, the identity of the interaction partner may also play a role in evoking facial behavior [9]. The social impact on facial expression in positive emotional settings was further investigated by other researchers [10] [11] with diverse conclusions, but with the main finding that facial expressions involve a complex interplay of three factors: emotional state, sociality of the situation and relationship with the co-viewer. Compared to the variety of studies on *positive* emotions, little work has yet been undertaken concerning the *negative* emotional setting. This is the main focus of this paper.

The methods used for measuring facial behavior in the previous studies can be categorized into three groups: (1) objective measurement using Electromyography (EMG) [12], (2) judgment in predefined categories [9], and (3) coding using facial behavior description systems, for example the Maximally Descriptive Facial Movement Coding System (MAX; [13]), the System for Identifying Affect Expression by Holistic Judgment (AFFEX; [14]) or the most widely used Facial Action Coding System (FACS; [15] [16]). The advantage of using the last type of measurement mentioned is that it is less obstructive and intrusive than EMG and more objective and detailed than pure experience-based judgment. Its disadvantage is also obvious that it is extremely labor-intensive. The Facial Expression Coding System (FACES; [1]) differs from FACS firstly for its theoretical background that the former bases on a *dimensional* model of emotion and the latter is often used to attribution of *discrete* emotions (e.g., basic emotions). Secondly, FACES involves merely judgment of facial expression on a valence dimension (the second category mentioned) whereas FACS describes the facial muscle movement in detail. The former largely shortens the training time (100 into 20 hours) and coding time (1h into 15min for one-minute video) and was effective enough for the hypothesis. Thus FACES was adopted in this study.

Method

Participants. Each participant was asked to bring a same-sex friend to the experiment, whom the participant had known for at least 3 months. A total of 48 German male participants (24 pairs; M= 25.99 years old, SD = 5.23) took part in the experiments. Half of them had been randomly assigned to positive emotional settings, which belongs to a part of the study beyond the scope of this paper. Participants were unaware of the real motivation for the study and were told that the purpose of the experiment was to investigate the relationship between emotion and visual attention. Each participant received 10 Euros as compensation. One additional male participant was recruited to play the role of a stranger, who had never met any of the paired participants. He was trained to behave in the same manner across all experiments.

Materials. Three sadness-inducing film clips were selected from validated film databases [17]. A pre-test was conducted with German participants. Results revealed a satisfactory capacity of the clips in inducing comparable high levels of distress without eliciting non-target emotions. To assess the participants' emotion, self-reports were obtained with the Differential Emotional Scale (DES) [18] consisting of 30 items related to 10 basic emotions.

Setting and Equipment. Two identical 15-inch notebooks with built-in cameras were placed at opposite sides of a conference table to play film clips and covertly record the facial expressions of participants. Thus participants were given the chance to look at each other's face and were aware of watching the same film clip with more exposure to the co-viewer's facial expression than when sitting besides each other. The experiments were programmed with OpenSesame [19] to guarantee synchronized play on the two screens under conditions of social viewing. Two external loudspeakers were used for playing the sound. Another 15-inch notebook equipped with headphones was placed in an adjacent room to run a visual attention task (see below).

Procedures. Each participant completed 3 film viewing sessions involving different social context (alone, with stranger, with friend) in a counter-balanced order with 2 visual task sessions between the film sessions as intersessions. Participants were instructed not to talk when watching the film clips. After each film-viewing session, participants rated their emotional experience with DES. At the very beginning of the whole experiment, participants viewed a neutral film as a baseline. The difficulty of the visual attention task was controlled to an acceptable level to avoid emotional artifacts and to allow the aroused emotions from last film session to diminish.

Experimental Hypothesis. According to Fridlund [2] sad faces in negative emotional settings may manifest the motive to seek for "succor" or request for "restitution". Thus it was hypothesized that the expressiveness of the facial display would vary as a function of sociality, namely the intensity of sad faces should show an increment between participants having a co-viewer (stranger or friend) and those viewing alone. Also, based on the findings that the identity of the co-viewer may have impact on the facial behavior of the subject [9], the facial behavior was expected to be more expressive in the friend-condition than in the stranger-condition.

Coding Facial Behavior

The video recordings were coded using Facial Expression Coding System (FACES; [1]). Raters were trained to detect facial behaviors that are emotion-related rather than serving other functions. They were asked to judge its expressed emotion valence and intensity and to note the respective duration. Previous validity studies indicate that FACES ratings are related in predictable ways to EMFACS, facial muscle activity (EMG), reports of experienced emotion and other psychophysiological measures [20].

Two graduate students were trained to code the facial expressions of the dummy subject (stranger) during each video clip using Noldus Observer XT 7.0. Raters coded with the same predefined coding scheme:

1) Facial expression: valence (positive or negative); duration (start time, end time); intensity (4-point rating scale).
2) Communicative behavior: glance at co-viewer (count).
3) Overall assessment of the whole video: dominant emotional expression (categorical: one of the six basic emotions); overall expressiveness (5-point rating scale). The overall assessment made by coders was only used for a manipulation check rather than qualitative or quantitative behavioral analysis.

Ratings were made with the volume turned off in order to prevent contamination of ratings due to speech content. The coding process was largely simplified due to the statistical module provided producing descriptive statistics of the observations. The software generated a summary of the frequency, duration and intensity of observed facial behavior. The inter-rater agreement was calculated using the intraclass correlation coefficient (ICC; Model 3.1) [21]. With satisfactory inter-rater agreement in the training phase, the raters continued to rate the videotapes of real subjects and achieved adequate inter-rater reliability ranging from 0.64 to 0.94. The scores on frequency, duration and intensity of positive or negative expressions were averaged from the two rater scores.

Conclusion

As an alternative to FACS coding, FACES has offered an effective and efficient means of assessing facial behavior based on dimensional emotional theories and required relatively less training and coding time. Together with the Observer XT, the coding process has been simplified into an affordable range, with all the data digitalized and prepared for further analysis.

Generally, an audience effect was found on positive expression in spite of the negative valence of the film stimulus, whereas negative expression did not vary as a function of social context (see Figures 1, 2 below). The facial expression and the self-reported emotion were not correlated under any social context. It may be inferred that in a social interactive context, participants seem more likely to “prevent” themselves or others from being affected by negative emotions than to “share” a negative emotional experience with the social interactant. Participants tended to report less experienced distress with the presence of a friend than when in the other two conditions. This observation may be explainable by the facial feedback hypothesis [22] [23] that facial movement did influence the self-report of emotion or by an emotional contagion effect [24] from other-smiles. More detailed discussion of the findings will be included in the presentation.

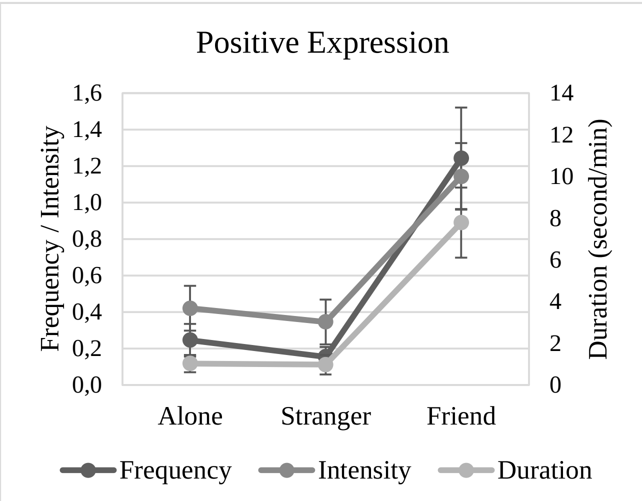


Figure 1. Mean frequency, duration and intensity of positive expression as a function of social context. Error bars represent standard error of the mean (SE).

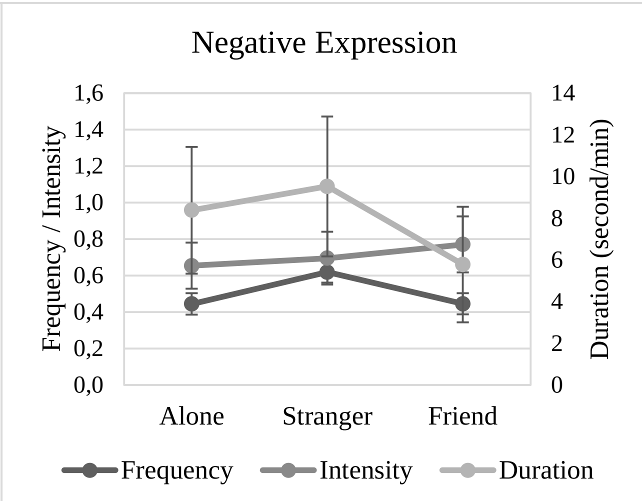


Figure 2. Mean frequency, duration and intensity of negative expression as a function of social context. Error bars represent standard error of the mean (SE).

References

[1] Kring, A.M., Sloan, D. (1991). *The Facial Expression Coding System (FACES): A users guide*. Unpublished Manuscript. Retrieved from http://socrates.berkeley.edu/%7Eakring/FACES%20manual.pdf [Last accessed on January 27, 2015]

[2] Fridlund, A.J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA, USA: Academic Press.

[3] Tomkins, S.S. (1962). *Affect imagery consciousness: Volume I, The positive affects*. New York, NY, USA: Springer.

[4] Ekman, P. (1972). Universals and cultural differences in facial expression of emotion. In J.K. Cole (Ed.), *Nebraska symposium on motivation*, p. 207-283. Lincoln, NE, USA: University of Nebraska Press.

[5] Izard, C.E. (1971). *The face of emotion*. New York, NY, USA: Appleton Century Crofts.

[6] Kraut, R.E., Johnston, R.E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, ***37***, 1539-1553.

[7] Chovil, N. (1991). Social determinants of facial displays. *Journal of Nonverbal Behavior*, ***15***, 141-154.

[8] Fridlund, A.J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, ***60***, 229-240.

[9] Wagner, H.L., Smith, J. (1991). Facial expression in the presence of friends and strangers. *Journal of Nonverbal Behavior*, ***15***, 201-215.

[10] Hess, U., Banse, R., Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology*, ***69***, 280-288.

[11] Jakobs, E., Manstead, A.S., Fischer, A.H. (1999). Social motives and emotional feelings as determinants of facial displays: The case of smiling. *Personality and Social Psychology Bulletin*, ***25***, 424-435.

[12] Fridlund, A.J., Cacioppo, J.T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, ***23***, 567-589.

[13] Izard, C.E. (1979). *The maximally discriminative facial movement coding system (MAX)*. Newark, DE, USA: University of Delaware Office of Instructional Technology.

[14] Izard, C.E., Dougherty, L.M. (1980). *System for identifying affect expressions by holistic judgment (AFFEX)*. Newark, DE, USA: Instructional Resources Center, University of Delaware.

[15] Ekman, P., Friesen, W.V. (1976). Measuring facial movement. *Journal of Environmental Psychology and Nonverbal Behavior*, ***1***, 56-75.

[16] Ekman, P., Friesen, W.V. (1978). *The Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychologists Press.

[17] Schaefer, A., Nils, F., Sanchez, X., Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, ***24***, 1153-1172.

[18] Izard, C.E., Dougherty, F.E., Bloxom, B.M., Kotsch, W.E. (1974). The differential emotions scale: A method of measuring the subjective experience of discrete emotions. Vanderbilt University.

[19] Mathôt, S., Schreij, D., Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social science. *Behavior Research Methods*, ***44***, 314-324.

[20] Kring, A.M., Sloan, D.M. (2007). The facial expression coding system (FACES): Development, validation, and utility. *Psychological Assessment*, ***19***, 210-224.

[21] Shrout, P.E., Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, ***2***, 420-428.

[22] James, W. (1890). *The principles of psychology*. New York, NY, USA: Dover Publications.

[23] Laird, J.D. (1984). The real role of facial response in the experience of emotion: A reply to Tourangeau and Ellsworth, and others. *Journal of Personality and Social Psychology*, ***47***, 909-917.

[24] Hatfield, E., Cacioppo, J.T., Rapson, R.L. (1994). *Emotional Contagion*. New York, NY, USA: Cambridge University Press.

Monitoring the physiology of the creative process.

Y.D. van der Werf[1,2], J.B.F. van Erp[3]

[1] Dept Emotion and Cognition, Netherlands Institute for Neuroscience, an Institute of the Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands.
Y.van.der.werf@nin.knaw.nl
[2] Dept Anatomy and Neurosciences, VU University medical centre, Amsterdam, The Netherlands.
[3] Perceptual and Cognitive Systems, TNO, Soesterberg, The Netherlands
jan.vanerp@tno.nl

Most people would agree that creativity is what makes us unique in the animal kingdom. For some, art is the highest expression of the human spirit. Interestingly, we understand little of the processes that drive creativity in the artist, or the processes in people experiencing art. Are they the same or at least similar? Or do the production and the appreciation of art rely on different mechanisms? Both the artist and their audience often report a feeling of 'flow', which appears to involve a change in the perception of time and a modulation of emotional tone. Such a state bears at least a superficial resemblance to trance, meditation and certain drug-induced states. In addition, the content of the art may be reflected in the viewer's mental state. Subjective reports of experiencing romance, beauty, horror, disgust or pornography differ widely, while they can all be part of creating and of experiencing art.

We are currently investigating these 'productive' and 'receptive' aspects of art in a study using prose as the model art form. Both writing and reading involve prolonged periods of reasonably well-described feelings and behaviours, with relatively little movement that might interfere with measurement, unlike, for example, the performing arts.

We have obtained detailed and multimodal measurements of Dutch author Arnon Grunberg while writing his new novel in a period of two weeks. The measurements included ambulatory recording of electrical brain activity (28-electrode electroencephalography - EEG), heartbeat (electrocardiography - ECG) and bodily processes reflecting arousal and emotion (galvanic skin response - GSR), under continuous recording of the writer's facial expressions. Each day, we performed a standardized measurement of basic emotions that we induced in the writer by showing him emotional images and asking him to write a portion of text according to the emotion portrayed; after this standardized measure, we commenced the recording of the author writing his novel. With these recordings in one individual, we aim to obtain a proof-of-principle of the multimodal recordings and will attempt dissociating the different emotions in the standardized measurements and the 'free writing' epochs, using novel and state-of-the-art methods (a.o. template matching algorithms). Using these methods, we will perform the second phase of the project, in which we will study members of the public reading the novel in a controlled laboratory setting (n=~100). A third phase of the project involves recording brain activation in several thousand readers, using novel hard-and software that will be developed as part of this project in a collaboration with the industry.

Future efforts will involve broadening the project to include other art forms alongside reading and writing (for example, composing and listening, cooking and eating). We foresee that the technical developments, in terms of software and analysis techniques, will impact on fields not necessarily related to the arts, such as sports, ergonomics, healthcare and education.

Continuous affect state annotation using a joystick-based user interface

J. Antony[1], K. Sharma[1*], C. Castellini[1], Egon L. van den Broek[2], C. Borst[1]

[1]Robotics and Mechatronics Center, German Aerospace Center, Wessling, Germany.
[2]Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands.
[*] karan.sharma@dlr.de

Abstract. Ongoing research at the DLR (German Aerospace Center) aims to employ affective computing techniques to ascertain the emotional states of users in motion simulators. In this work, a novel user feedback interface employing a joystick to acquire subjective evaluation of the affective experience is presented. This interface allows the subjects to continuously annotate their affect states, elicited in this scenario by watching video clips. Several physiological parameters (e.g. heart rate, electrodermal activity, respiration rate, etc.) were acquired during the viewing session. A statistical analysis is presented, which shows expected patterns in data that validate the design and methodology of the experiment and lay the groundwork for further experiments to be undertaken at the DLR.

Introduction

The work presented here aims to ascertain the affective experience of a human subject in a motion simulator, namely the DLR-Robotic Motion Simulator (DLR-RMS) [1]. Russell defines *core affect* as *'the neurophysiological state that is consciously accessible as a simple, non-reflective feeling that is an integral blend of hedonic (pleasure-displeasure) and arousal (sleepy-activated) values'* [2]. The study presented here focuses on 4 representative and contrasting emotion classes that can presumably be elicited during the course of a simulation, namely *thrill/amusement, fear, boredom and calmness*. These emotion classes are characterised by the different *core affect* (i.e. valence-arousal levels) states associated with them [1]. With the medium-term goal of quantifying in real-time and in a continuous fashion the affective state of a subject during one of the aforementioned motion simulations, we designed an initial ground-based experiment in which video clips were administered to 30 human subjects; while viewing the clips, each subject was required to annotate his/her affective state using a joystick, and several kinds of biological signals were recorded. Some of the important design aspects of this experiment are explained in the following sections.

Use of Videos for Affect Elicitation

Video stimuli have been used in emotion studies for eliciting specific and intense emotional responses in a relatively short period of time [3,4]. Video stimuli are dynamic, encompassing visual and auditory cues which lead to greater involvement of the participant in the emotion experience [3]. Also, carefully selected video stimuli can be used for the elicitation of the intended affect states (both positive and negative) in sufficient intensity and duration without involving any harm or ethical violation [4]. Although an affective experience on a motion simulator also involves motion, the use of video stimuli offers a similar affective experience to a simulation and is therefore used in this initial investigation.

Use of Biosignals for determining the Affect State

During the course of the experiment, several physiological parameters of users were measured by means of different bio-sensors. These include: electrocardiogram (ECG) and the blood volume pulse (BVP) sensors to measure the cardiac activity, an electrodermal activity (EDA) sensor, a respiration belt sensor to acquire the respiration rate, a skin temperature sensor and three surface electromyography (sEMG) sensors to measure the muscular activity in the *Corrugator Supercilii*, the *Zygomaticus Major* and the *Trapezius* muscles. These biosignals are indicative of the underlying affect states [5,6,7]. In the context of affect state estimation, biosignals have the desirable properties of being easily acquired, immune to voluntary manipulation or masking, generally independent of age and gender, etc. [7].

Annotation Methods

Traditional self-reporting/assessment techniques, allowing participants to state their affect/emotional states, are post-stimuli questionnaires such as, e.g. Likert scales [4,8]; however, most of these methods don't allow for continuous self-reporting and therefore the flow of the experiment has to be interrupted to allow the participant to report their experience [9]. To overcome this limitation, self-reporting HMIs using dials [9] or computer mice [9] have been used; along similar lines, we developed a joystick-based user interface (UI) that allows for continuous annotation of affect experiences by positioning a pointer in a 2-D space. The motivation behind using a 2-D annotation space is that according to Lang [9], "all emotions can be located in a 2-D space, as co-ordinates of affective valence and arousal". Therefore, by allowing the user to report their perceived valence and arousal levels instead of discrete emotional states, the interface allows the users to also report mixed emotional experiences. Another advantage that a joystick offers over other methods is that it is intuitive to use and enables the user to annotate his/her affect experience at the same time as they are elicited. This helps in reducing the measurement errors in emotion assessment arising from delays in reporting, fading or distortion of affect state experiences during recall and errors due to distraction of the participant from the emotion stimuli [4,8].

Setup Description: The Annotation Interface

Since this paper focuses on the annotation method used for the experiment, a detailed explanation of the experimental methodology and procedure is not presented here. In the following sub-sections, a concise overview of different aspects of the annotation method is presented.

Design of the UI: The interface design is in essence based on the 2-D valence-arousal affect model (see Figure 1) [12]. One difference is the addition of Self-Assessment Manikin (SAM) [12,13] to the co-ordinate axis of the interface. The manikin figures depict different valence (on X-axis) and arousal (on Y-axis) levels and are arranged on the co-ordinate axes in an ascending order of these levels. A red pointer, seen at the position (5,5) in Figure 1, represents the rest position of the joystick and is defined as the neutral affect position. The UI is displayed on the upper right corner of the LCD monitor (see Figure 2). This allows the user to view the video-clip and annotate his affect state simultaneously.

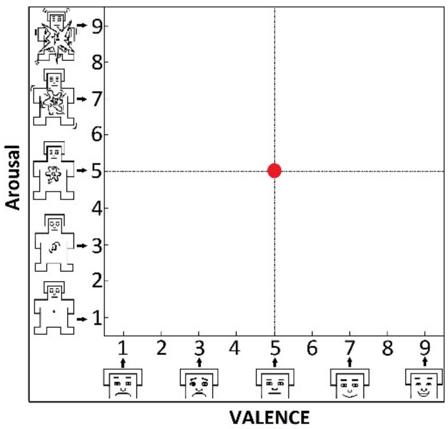


Figure 1. The annotation UI.

Figure 2. The annotation UI embedded in a video.

Annotation Procedure: While viewing a video-clip, the user undergoes an affective experience resulting in a change in his valence and arousal levels. The user annotates his perceived experience by positioning the red pointer in the appropriate region of the UI. When the elicited affect changes, the user is supposed to then position the pointer in the region that best characterizes his/her perceived affect state.

Training Procedure: A few days before the test, the subjects were provided a document with an overview of the annotation interface. On the day of the test, before commencing the session, they were again given a short re-introduction to the valence-arousal model and the annotation procedure. They were instructed not to annotate the affective content of the video, but rather provide ratings based on their perception of their affect state. After this, they were given a short demo of the UI and were allowed to get a feel of the system. This was followed by a practice session containing 5 video-clips of 1 minute duration. During this session, the users annotated their affect states while watching the videos and were allowed to ask questions (pertaining to the annotation process) at the end of each video-clip. After the training session, the main experiment was started.

Data Analysis and Results

Figure 3 shows the affect state annotation values of a typical subject for the entire duration of the test. The *x* and *y* co-ordinate values of joystick data (see Figure 1) provide the valence and arousal levels, respectively. The data are colour coded (and labelled) based on the emotion that the video clip was expected to elicit. The emotion label associated with each video was determined through an initial evaluation, undertaken by a different set of participants. Except the first two videos in the sequence (marked as ‘Calmness’ and ‘Neutral’ in Figure 3) and the last video (marked as ‘Neutral’ in the Figure 3), the order of all the other videos was randomly shuffled and every participant viewed a different sequence. Also, no two videos eliciting the same type of emotional response would ever be shown one after another, and each video was followed by a blue screen (duration: 2 minutes, marked at ‘Neutral’ in Figure 3) to isolate the impact of each video on the affect state of the participant. The total number of videos in a sequence was 18, i.e. 8 videos for emotion elicitation (2 for each emotion label), 9 blue screens and 1 ‘calming’ video at the start of the sequence. After the test, the users provided feedback on the usability of the annotation system by filling out a System Usability Scale (SUS) survey form. An analysis of the results of this survey is not presented here as it’s beyond the scope of this publication.

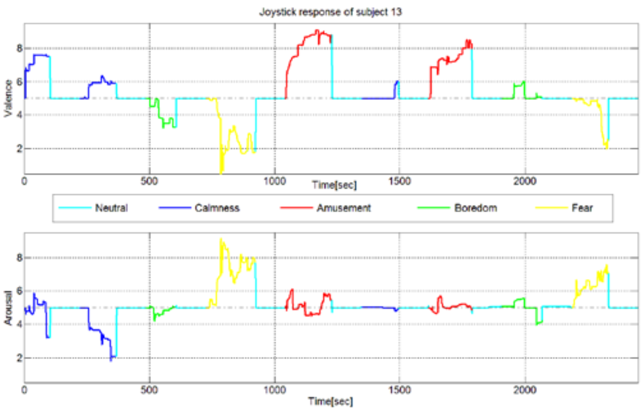


Figure 3. Annotation values for subject 13.

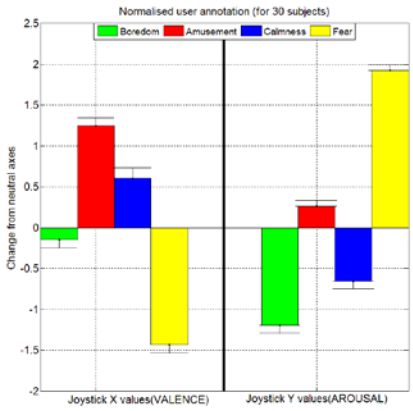


Figure 4. Normalised annotation values for all subjects.

A ‘scary’ video is expected to elicit ‘fear’ in the participants, which is characterized by high arousal and low valence in the valence-arousal model of affect. From the annotation data for the two horror clips (labelled as ‘fear’ and represented by the yellow segments in Figure 3 & 4) used in this test, it can be seen that the response of subject 13 to these videos is categorized by low valence and high arousal values. In the valence-arousal affect model, the other type of video segments eliciting boredom, amusement and calmness are classified by: low valence-low arousal, high valence-high arousal and intermediate valence-low arousal levels, respectively. The response of subject 13 (see Figure 3) is along the same lines.

Figure 4 shows the mean valence and arousal values and the standard error of means (SEM) for all 30 subjects, across different video labels; valence and arousal values from every participant were first averaged across every video clip, then these averaged values were normalised. From the normalised dataset, the mean and the SEM values for each video type, across all 30 subjects were calculated. The *x*-axis in this figure is equivalent to neutral (5,5) position in the UI (see Figure 1). Therefore, negative values in this figure signify low valence and arousal values (from 1 to 5) in the UI. From this plot, it can be observed, e.g. that ‘fear’ eliciting video clips are annotated as high arousal and low valence events by all users. Similarly, ‘calm’ videos are represented by low arousal and intermediate valence.

Conclusion

Using a joystick based annotation UI offers several advantages over conventional self-reporting methods. For example, through the use of a joystick based UI, the video clips are continuously annotated. Therefore, in the case of 'fear' eliciting videos, we can also explore the 'build up' phase for an affect state, which is of interest [cf. 14,15]. Secondly, any errors in labelling originating from temporal latency are also avoided [cf. 16]. Based on the results presented above, we state that the proposed annotation system is a viable and favourable alternative to the conventional self-reporting techniques.

References

[1] Bellmann, T., Heindl, J., Hellener, M., Kuchar, R., Sharma, K., Hirzinger, G. (2011). The DLR Robot Motion Simulator Part I: Design and Setup. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA 2011)* (Shanghai, China, May 9-13), p. 4694 –4701.

[2] Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review* ***110****(1)***,** 145-172.

[3] Gross, J. J., Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion* ***9****(1)***,** 87-108.

[4] Rottenberg, J., Ray, R.D., Gross, J.J. (2007). Emotion elicitation using films. In J.A. Coan, J.J.B. Allen (Eds.), *Handbook of emotion elicitation and assessment, Chapter 1*, p. 9-28. New York, NY, USA: Oxford University Press, Inc.

[5] Picard, R.W. (1995). *Affective computing*. Technical Report 321, M.I.T Media Laboratory Perceptual Computing Section, Cambridge, MA, USA.

[6] Liu, C., Rani, P., Sarkar, N. (2006). Human-robot interaction using affective cues. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN06)* (Hatfield, UK, September 6-8), p. 285–290.

[7] Van den Broek, E.L. (2011). *Affective Signal Processing (ASP): Unravelling the mystery of emotions.* CTIT Ph.D.-thesis series No. 11-204, Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.

[8] Levenson, R.W. (1988). Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In H.L. Wagner (Ed.), *Social psychophysiology and emotion: Theory and clinical applications, Chapter 2*, 17-42. Chichester, West Sussex, Great Britain: John Wiley & Sons Ltd.

[9] Ruef, A.M. and Levenson, R.W. (2007). Continuous measurement of emotion: The affect rating dial. In J.A. Coan, J.J.B. Allen (Eds.), *Handbook of emotion elicitation and assessment, Chapter 17*, p. 286-297. New York, NY, USA: Oxford University Press, Inc.

[10] Nagel, F., Kopiez, R., Grewe, O., Altenmüller, E. (2007). Emujoy: Software for continuous measurement of perceived emotions in music. *Behavior Research Methods,* ***39****(2)*, 283-290.

[11] Lang, P.J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist,* ***50****(5)***,** 372-385.

[12] Van den Broek, E.L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing,* ***17****(1)*, 53-67.

[13] Bradley, M.M. and Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry,* ***25****(1)***,** 49-59.

[14] Van den Broek, E.L., Westerink, J.H.D.M. (2009). Considerations for emotion-aware consumer products. *Applied Ergonomics,* ***40****(6)*, 1055-1064.

[15] Van den Broek, E.L., Lisý, V., Janssen, J.H., Westerink, J.H.D.M. , Schut, M.H., Tuinenbreijer, K. (2010). Affective Man-Machine Interface: Unveiling human emotions through biosignals. In A. Fred, J. Filipe, H. Gamboa (Eds.), *Biomedical Engineering Systems and Technologies,* p. 21-47., Berlin/Heidelberg, Germany: Springer-Verlag.

[16] van den Broek, E.L., van der Sluis, F., Dijkstra, T. (2013). Cross validation of bi-modal health-related stress assessment. *Personal and Ubiquitous Computing,* ***17****(2)*, 215-227.

Environmental Enrichment and Data Quality: Results of a Global Survey

P. Hawkins

Research Animals Department, Royal Society for the Prevention of Cruelty to Animals (RSPCA), Southwater, U.K. penny.hawkins@rspca.org.uk

Introduction

It is widely accepted within the scientific community that 'better welfare equals better science', and it is also commonly acknowledged that environmental enrichment improves animal welfare. For example, recent versions of legislation, guidelines and codes of practice relating to laboratory animal use emphasise the importance of a stimulating environment to encourage appropriate natural behaviours [1-3], including group housing for social animals and environmental enrichment such as nesting material, refuges and chew blocks for rodents. However, there are still inconsistencies in the provision of enrichment. Reasons for this were recently reviewed in an article in the *Enrichment Record*, including the belief that experimental variability will increase, or a confound will be introduced, affecting data quality [4]. There are three main areas of concern relating to this: (i) whether the validity of the science within an individual project will be affected; (ii) whether the data will still be comparable with those obtained from studies conducted without enrichment; and (iii) whether greater variability will require more animals for significant results.

The article discussed the scientific, practical and ethical bases for each of these, suggesting action points to support the provision of enrichment (and encourage wider reporting of husbandry refinements) [4]. It also asked readers to participate in a survey, drawn up by the RSPCA (a UK scientific animal welfare organisation) and GR8 (a global education and training organisation), aiming to explore views and beliefs about the effects of enrichment on data quality. This ran for seven weeks, ending on 21 February 2014, and was promoted in the *Enrichment Record* article and via online forums such as Compmed, LAREF and laboratory animal science-related user groups on LinkedIn, including the Measuring Behavior group. The aim was to encourage a wide range of respondents, holding a variety of attitudes towards enrichment and beliefs about its effects on data quality [5].

Survey Results

A total of 343 people responded, including 112 researchers, 112 animal technologists, 85 veterinarians, 74 members of ethical and animal care and use committees and 9 regulators (the total number of responses was 473, as respondents could tick more than one role). Most (182 = 53 %) were from academic establishments or universities, 51 (15 %) worked in pharmaceutical establishments, 41 (12 %) in medical or veterinary research institutes, 26 (8 %) in government agencies and 25 (7 %) in Contract Research Organisations. The majority of respondents were located in north America (175 = 51 %) or Europe (128 = 37 %), with 20 (6 %) from Canada and small numbers of others located in most areas of the world apart from Oceania and sub-Saharan Africa. For full survey results, see [5].

When asked whether they regarded environmental enrichment as a basic necessity for animals, 336 people responded 'yes' and 5 answered 'no'. This level of acceptance was also reflected in the responses to the question: '*In your view, is there adequate scientific evidence that animals benefit from enrichment?*' Most respondents (275) agreed with the option '*Yes – I can accept the methodologies used to evaluate this and their conclusions*', whereas 53 agreed with '*For some species, but not necessarily all*'. Over 95% of all 343 respondents thus agreed that there is scientific evidence that enrichment benefits at least some species, with 80 % of all respondents choosing the option that did not include a caveat relating to different species. Just 10 people (3 %) were not convinced.

The results of the survey should therefore be interpreted as representing the views of people who, in general, regard enrichment as a basic necessity for animals and accept the scientific evidence that animals

benefit from its provision. Other responses indicated that the majority of respondents believe that enrichment has a positive effect on data quality, and work in an environment that is broadly supportive of providing enrichment for animals. They have a good level of awareness of the literature on the effects of enrichment, are prepared to encourage others to provide it, and recognise that it is relevant to the science and should be reported in publications.

However, the responses identified four issues of concern:

- Withholding environmental enrichment on the basis of untested assumptions about its impact on data quality (18 % of 243 respondents who felt the relevant question applied to them)
- A lack of motivation to challenge colleagues who do not provide enrichment, or a lack of support when raising the issue (15 % of 190 respondents to whom this applied)
- Perceptions that enrichment is irrelevant to materials and methods sections of publications (15 % of 135 people)
- Lack of a properly defined system (either via designated individual(s) or an ethical or animal care and use committee) for researching, retrieving and assessing new information on enrichment (see Figure 1)

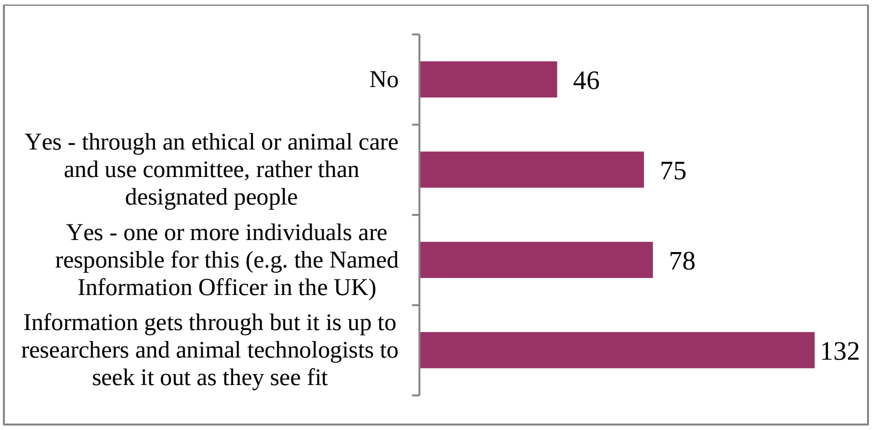


Figure 1. Responses to the question: '*Does your facility have a system for researching, retrieving and assessing new information on enrichment?*

All of the above issues arose even among people who were already engaging with welfare issues, motivated to provide enrichment and working in facilities that in general had a good 'culture of care', so they may well be more widespread among the wider population. The conclusions below are set out on that basis.

Discussion and Conclusions

Enrichment should be provided unless there is sound scientific or veterinary justification. This relates to concern (i) in the Introduction. It is essential to evaluate the impact of enrichment on data quality if there are any doubts about this, especially in the field of behavioural research where animal housing, husbandry and care can have significant effects on the parameters under investigation [6]. To give just two examples, enrichment has been reported to prevent cognitive impairment in rats with epilepsy [7] and enhance episodic-like memory persistence in an object recognition paradigm in mice [8]. From a translatability aspect, it could be argued that the validity of the results in studies such as these may be enhanced, as humans also live in an 'enriched' environment. In that case, any concerns about comparability with data obtained from studies without enrichment (concern (ii) in the Introduction) would not in themselves justify continuing to withhold enrichment, as the science would be improved.

Pilot studies can provide an informed basis for decision making about enrichment, by quantifying any potential effects and (if necessary) indicating necessary adjustments to experimental design, and data analysis, to accommodate any differences. In some cases this might involve an increase in animal numbers, which presents a conflict between reduction and refinement (issue (iii) in the Introduction). However, the consensus view is that reducing suffering to each individual – and improving welfare – should usually be the priority, even if more animals are used (just 5 % of survey respondents believed that reduction always came first). There will be further justification for increasing numbers if enrichment is found to improve the science as well as animal welfare.

People who are supportive of providing enrichment should increase their efforts to communicate about it and 'set a good example', both within their faciliities and in their publications. Colleagues and journal editors should support them in this. Most respondents who encouraged colleagues to provide enrichment were successful, which is very positive, with a reported success:failure rate of around 3:1. However, it was disappointing to see that others were not motivated or did not feel they would be supported. This is largely a cultural issue that can be addressed within individual facilities, and it would be a good topic for an institutional ethical or animal care and use committee to discuss with staff.

Communication about animal housing, husbandry and care, including enrichment, within materials and methods sections of publications could also be improved. This is essential information to help disseminate and encourage good practice with respect to refinement, and to ensure all potential variables have been described so that the results can be adequately interpreted, compared with those from other facilities and replicated if necessary [9]. However, there are currently serious issues with inadequate levels of reporting in biomedical research papers [10-12] and a number of reporting guidelines have been drawn up to help to address this [e.g. 10,13]. Authors are beginning to use these, either in their entirety or adapted to individual studies, but reporting guidelines could still be used more widely – both at the protocol design stage and when preparing papers for publication [12]. It is noteworthy that Measuring Behavior has taken this responsibility seriously within its guidance for authors [14].

Every establishment should have a system for ensuring that new information on enrichment is researched, retrieved, assessed and implemented where appropriate, to facilitate better science and better welfare. Otherwise, publications and other communications on enrichment can easily be overlooked, for a number of reasons including lack of awareness of relevant journals, discussion fora and meetings that feature enrichment, or an establishment culture that does not recognise the value of regularly reviewing new information on refinement including enrichment. The role of ensuring an adequate flow of information can be undertaken by one or more designated individual(s), such as the Named Information Officer in the UK, or an institutional committee. NB ensuring that staff have access to species-specific information is also a legal requirement within the European Union [1] and recognised as good practice in other guidelines [2].

Helpful information and guidance relevant to all of the above can be found in *Guiding Principles for Behavioural Laboratory Animal Science*, which was supported and funded by the British Association for Psychopharmacology (BAP), British Neuroscience Association (BNA), the ESSWAP Foundation and the UK Laboratory Animal Science Association (LASA) [15].

References

[1] European Commission (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the Protection of Animals Used for Scientific Purposes. *Official Journal of the European Union L*, ***276***, 33-79.

[2] National Research Council (2011). *Guide for the Care and Use of Laboratory Animals: Eighth Edition*. Washington, DC, USA: The National Academies Press.

[3] CIOMS/ICLAS (Council for International Organization of Medical Sciences, The International Council for Laboratory Animal Science) (2012). *International Guiding Principles for Biomedical Research Involving Animals.* http://iclas.org/wp-content/uploads/2013/03/CIOMS-ICLAS-Principles-Final1.pdf [Last accessed on 11 March 2014]

[4] Hawkins, P. (2014). Facts and Demonstrations: Exploring the Effects of Enrichment on Data Quality. *The Enrichment Record* **Winter 2014**, 12-21. http://enrichmentrecord.com/wp-content/uploads/2014/01/FACTS-AND-DEMO.pdf [Last accessed on 11 March 2014]

[5] Hawkins, P. (2014). Results of the GR8/RSPCA Global Survey on the Effects of Enrichment on Data Quality. *The Enrichment Record.* http://enrichmentrecord.com/wp-content/uploads/2014/04/RESULTS-OF-GR8.pdf [Last accessed on 28 May 2014]

[6] Simpson, J., Kelly, J.P. (2011). The Impact of Environmental Enrichment in Laboratory Rats – Behavioural and Neurochemical Aspects. *Behavioural Brain Research,* ***222***, 246-264.

[7] Fares, R.P., Belmeguenai, A., Sanchez, P.E., Kouchi, H.Y., Bodennec, J., Morales, A., Georges, B., Bonnet, C., Bouvard, S., Sloviter, R.S., Bezin, L. (2013). Standardized environmental enrichment supports enhanced brain plasticity in healthy rats and prevents cognitive impairment in epileptic rats. *PLoS ONE,* ***8****(1)*, e53888.

[8] Leger, M., Quiedeville, A., Paizanis, E., Natkunarajah, S., Freret, T., Boulouard, M., Schumann-Bard, P. (2012). Environmental enrichment enhances episodic-like memory in association with a modified neuronal activation profile in adult mice. *PLoS ONE,* ***7****(10)*, e48043.

[9] Osborne, N.J. (in press). Where Research Integrity and Scientific Reporting Converge, Animal Welfare Stands to Benefit. In: *Proceedings of the 3rd World Conference on Research Integrity*, 5-8 May 2013, Montréal, Canada. http://www.wcri2013.org/ [Last accessed on 11 March 2014]

[10] ILAR (2011). *Guidance for the Description of Animal Research in Scientific Publications.* Washington, DC: The National Academies Press. http://www.nap.edu/catalog.php?record_id=13241 [Last accessed on 7 March 2014]

[11] Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M.F.W., Cuthill, I.C., Fry, D., Hutton, J., Altman, D.G. (2009). Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals. *PLoS ONE,* ***4****(11)*, e7824.

[12] Baker, D., Lidster, K., Sottomayor, A., Amor, S. (2014). Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies. *PLoS Biology,* ***12****(1)*, e1001756.

[13] Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G. (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLoS Biology,* ***8****(6)*, e1000412.

[14] http://www.measuringbehavior.org/mb2014/measuring-behavior-conferences-and-animal-welfare [Last accessed on 12 March 2014]

[15] BAP, BNA, ESSWAP, LASA (2013). *Guiding Principles for Behavioural Laboratory Animal Science* (1st ed.), November, 61, available at http://www.esswap.org/guiding-principles-for-behavioural-laboratory-animal-science.html [Last accessed on 6 March 2014]

Measuring positive emotions in dairy cows using ear postures

H.S. Proctor

World Animal Protection, International, London, UK

helenproctor@worldanimalprotection.org

The importance of positive emotions

Animal welfare science is increasingly concerned with the promotion of positive emotions in animals [1,2]. Good animal welfare is now considered to not only require the absence of negative states, but also the active promotion and experience of positive states and emotions such as pleasure and joy [3,4]. Despite this, there is still very little research seeking to understand and measure positive states in animals [4]. In a recent systematic review of the literature Proctor *et al*. found there to be a significant bias towards the measurement of negative emotions, with very little research being conducted on positive states in animals [4]. If we are to promote positive states in animals we need to know much more about them, including how to measure them reliably [5].

Measuring emotions

Emotions are subjective entities, personal to the individual and therefore difficult to interpret and measure. The difficulties are only increased when we deal with animals, who do not share a common language with humans. However, this difficulty also applies to humans, as not all are capable of speech, and all have the potential to consciously or unconsciously misreport their emotional state.

Emotions are inbuilt in animals as they are in humans, and are essential in enabling animals to communicate with others, interpret situations correctly, and facilitate responses [1,6]. Emotions have significant adaptive value and their importance in relation to communication provides a useful tool for measuring them.

Ear postures in cattle

Ruminants have highly developed muscles around their ears, enabling them to independently move their ears into various postures [7]. Studies have investigated whether these postures are indicative of emotional state in both pigs and sheep [7,8], but none to date have looked at cattle. In this study we examined whether ear postures in dairy cows were reliable measures of a low arousal, positive emotional state.

Subjects and housing

Data collection was carried out between October and December, 2013, on 13 habituated dairy cows housed at Bolton Park Farm, Hertfordshire, UK. The cows had all been brought indoors for the winter and did not have access to grazing. The cows used in the study were separated daily by the farm staff from the main herd into two indoor pens following each morning milking. They then re-joined the main herd following the afternoon milking. The cows were otherwise maintained under standard feeding and handling procedures.

Positive stimulus

We imitated allogrooming in dairy cows, an experience considered to induce a positive,low arousal state [9,10], by performing stroking on the head, neck and withers of 13 habituated cows. The areas stroked were identified as being positively received and preferred in previous studies [9,11]. The positive effects of tactile contact have already been used to improve interactions between animals and stock-people [11,12]. Stroking of dairy cows and heifers has also been shown to reduce their fear of humans [13], reduce their cortisol levels [14], and lower their heart rate during various procedures [15,16]. Furthermore, both in this study and in previous studies, cows have been found to pursue the retreated stroker following a stroking bout and accept a second bout. This suggests that the cows were not just curious but actually found the experience pleasurable and wanted to prolong it [17].

Experimental procedure

We conducted 381 focals, each comprised of three, 5 minute conditions; pre-stroking, stroking, and post-stroking. To begin, the focal cow was randomly selected and the relevant details recorded (e.g. start time, cow number, and the side upon which the stroking and filming were to be performed). Each focal involved three researchers. Researcher one was responsible for filming the cow's ear on the same side which was being stroked. To do this they used a Sony HDRXR160EB Handycam Camcorder, mounted on a monopod, and stood at a distance of 2-4 feet from the cow. Researcher two used continuous sampling to record the focal cow's behaviour using a pre-defined ethogram. Frequencies were recorded for 15 of the behaviours listed on the ethogram, and durations were recorded for the behaviours lying and ruminating. The third person, the stroker, remained present during the entire focal, in order to rule out any effects on the ear postures resulting from a change in the number of people present.

The pre-stroking session lasted for 5 minutes. After this the stroker began the 5 minute stroking bout. The stroker wore thick canvas gloves and concentrated on the preferred regions of the cow's head, neck and withers at a rate of 40-60 strokes per minute to replicate the speed with which a cow would receive allogrooming [11]. After 5 minutes, the stroker stopped stroking and walked away from the cow. Researchers one and two continued to film and record the behavioural observations for the following 5 minutes (post-stroking session).

In order to ensure that the cow found the experience positive, any focal in which the cow moved away from the stroker was aborted. Furthermore, if the cow began to feed at any point during the 15 minutes, the focal was also aborted, as there was a risk that feeding could be an additional positive experience for the cow, and therefore impact the effects of the stroking. If the camera's view of the cow's ear was obstructed for longer than 30 consecutive seconds the focal was also aborted.

To control for the effect of activity levels on the cow's arousal levels we performed focals both when the cows were lying down, and when they were standing. We performed 309 full focals with lying cows, and 72 with standing cows. The total 381 focals were performed evenly across the group, and each cow was stroked equally on their left and right side to control for any effect of cerebral lateralisation.

Ear posture identification

We identified four unique ear postures performed by the cows; two alert, and two relaxed postures. Ear posture one (EP1) was characterised by the ear being held upright upon the cow's head, whereas ear posture three (EP3) appeared to be a more focussed posture with the pinna directed forwards, usually in the direction of a specific noise or activity. Ear posture four (EP4) was the most relaxed posture with the pinna drooping towards the ground and the ear naturally falling perpendicular to the head. There was no muscular tension involved in this posture, and the ear was entirely relaxed. Ear posture two (EP2) was positioned between EP1 and EP4. The ear was not entirely relaxed, nor was it tensely held upwards or forwards as in EP1 and EP3. EP2 was characterised by a small degree of muscular tension in which the ear was held slightly back on the head and did not fall perpendicular to the head like EP4.

Predictions

We predicted that the alert postures, EP1 and EP3, would be performed for significantly less time during stroking compared with the non-stroking conditions; pre and post-stroking. We predicted that the opposite would occur for the relaxed postures which would be performed for significantly longer during stroking than during the non-stroking conditions. We also predicted that the number of ear posture changes would significantly drop during stroking compared with both of the non-stroking conditions.

Video analysis

Each video was analysed to determine both the number of ear posture changes and the duration of time each posture was performed. This was done separately for each condition in the focal; pre-stroking, stroking, and post-stroking.

Data analysis

Using IBM SPSS Statistics Version 22, we performed One-Way ANOVA repeated measures analyses to compare the duration of time spent in each posture during the pre-stroking, stroking and post-stroking conditions. Pairwise comparisons were used to identify the significant individual differences. To identify any effects of activity levels and ruminating behaviour we used Paired Samples T-Tests to compare the non-stroking data from each category (e.g. ruminating versus non-ruminating, and standing versus lying). We then performed One-Way ANOVA repeated measures analyses on the data from each of the categories separately; ruminating, non-ruminating, standing and lying. To compare the number of ear posture changes we performed a One-Way ANOVA repeated measures analysis, comparing each condition; pre-stroking, stroking and post-stroking. To determine whether there were any lateralisation effects when stroking the cows on either the left or right side, we analysed the left and right focals using the Paired Samples T-Test. All assumptions were met for both the One-Way ANOVA repeated measures analyses and the Paired-Samples T-tests. When the Mauchly's test indicated that the assumption of sphericity had been violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity.

Results

We found that the two alert postures EP1 and EP3 were performed for significantly longer during the pre and post-stroking conditions than during stroking (EP1; $F(1.87, 671.09) = 241.22$, $p < 0.001$; EP3; $F(1.86, 668.87) = 39.09$, $p < 0.001$). The opposite effect was found for the relaxed ear postures EP2 and EP4, which were performed for significantly longer during stroking than during either the pre or post-stroking conditions (EP2; $F(1.95, 698.27) = 81.20$, $p < 0.001$, EP4; $F(1.65, 591.02) = 169.98$, $p < 0.001$). Furthermore, EP1 was performed for significantly less time, and EP2 for significantly more time during post-stroking compared with pre-stroking (EP1; $p < 0.001$, EP2; $p < 0.001$). We also found that the number of ear posture changes significantly increased during stroking, compared with both the pre and post-stroking conditions ($F(2, 718) = 17.89$, $p < 0.001$). All effects were significant, regardless of any effects of ruminating and activity levels. We found no lateralisation effects in the data when we compared the ear postures and number of changes performed during focals conducted on the left and right side of the cow.

Conclusion

This study showed that ear postures in dairy cattle are significantly affected by their experience of a positive, low arousal stimuli. Consequently, we suggest that this is reflective of their emotional state, and that ear postures may be used as a credible, and reliable measure of positive, low arousal emotional state. Further research is required to replicate this study using differing levels of arousal and different stimuli.

Finding reliable and objective measures of animal emotions is crucial if we are to improve animal welfare and ensure animals have a good life (Leliveld et al., 2013). Measures such as ear postures, offer a quick and easy tool for measuring and exploring the subjective minds of animals and can be used both in research and industry to improve the welfare of the billions of animals in our care. Studies such as this must continue to explore and develop measures of emotions so that we may apply our understanding of the emotional lives of animals in practice, and continue to improve the standards of animal welfare around the world.

Ethical statement

The experiment met with the Royal Veterinary College's ethical guidelines and the UK's Home Office advised that no review or license was required by them.

References

[1] Boissy, A., Manteuffel, G., Jensen, M. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behaviour, **92***, 375–397.

[2] Proctor, H. (2012). Animal Sentience: Where are we and where are we heading? *Animals, **2***, 628–639.

[3] Green, T. C., Mellor, D. (2011). Extending ideas about animal welfare assessment to include "quality of life" and related concepts. *New Zealand Veterinary Journal, **59***, 263–71.

[4] Proctor, H., Carder, G., Cornish, A. (2013). Searching for Animal Sentience: A Systematic Review of the Scientific Literature. *Animals, **3***, 882–906.

[5] Désiré, L., Boissy, A., Veissier, I. (2002). Emotions in farm animals: a new approach to animal welfare in applied ethology. *Behavioural Processes, **60***, 165–180.

[6] Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology, **288***, 1–20.

[7] Reefmann, N., Butikofer Kaszàs, F., Wechsler, B., Gygax, L. (2009). Ear and tail postures as indicators of emotional valence in sheep. *Applied Animal Behaviour Science, **118***, 199–207.

[8] Reimert, I., Bolhuis, J., Kemp, B., Rodenburg, T. (2012). Indicators of positive and negative emotions and emotional contagion in pigs. *Physiology & Behaviour, **109***, 42–50.

[9] Schmied, C., Waiblinger, S., Scharl, T., Leisch, F., Boivin, X. (2008). Stroking of different body regions by a human: Effects on behaviour and heart rate of dairy cows. *Applied Animal Behaviour Science, **109***, 25–38.

[10] Mendl, M., Burman, O.H.P., Paul, E. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society of Biological Sciences, **277***, 2895–904.

[11] Schmied, C., Boivin, X., Waiblinger, S. (2008). Stroking different body regions of dairy cows: effects on avoidance and approach behavior toward humans. *Journal of Dairy Science, **91***, 596–605.

[12] Windschnurer, I., Barth, K., Waiblinger, S. (2009). Can stroking during milking decrease avoidance distances of cows towards humans? *Animal Welfare, **18***, 507–513.

[13] Breuer, K., Hemsworth, P., Coleman, G. (2003). The effect of positive or negative handling on the behavioural and physiological responses of nonlactating heifers. *Applied Animal Behaviour Science, **84***, 3–22.

[14] Hemsworth, P., Barnett, J. (1989). The effects of handling by humans at calving and during milking on the behaviour and milk cortisol concentrations of primiparous dairy cows. *Applied Animal Behaviour Science, **22***, 313–326.

[15] Schmied, C., Boivin, X., Scala, S., Waiblinger, S. (2010). Effect of previous stroking on reactions to a veterinary procedure. Behaviour and heart rate of dairy cows. *Interaction Studies, **11***, 467–481.

[16] Waiblinger, S., Menke, C., Fölsch, D. (2003). Influences on the avoidance and approach behaviour of dairy cows towards humans on 35 farms. *Applied Animal Behaviour Science, **84***, 23–39.

[17] Bertenshaw, C., Rowlinson, P. (2008). Exploring heifers' perception of "positive" treatment through their motivation to pursue a retreated human. *Animal Welfare, **17***, 313–319.

Pigs suffering from injurious behaviours like flank biting and tail biting are more interested in manipulating a novel rope than uninjured control animals

M.B.M. Bracke[1*], K. Ettema[1]

[1] Livestock Research, Wageningen UR, Lelystad, the Netherlands
[*] Marc.bracke@wur.nl

Abstract

Injurious behaviours in pigs may involve persistent or forceful biting in specific body parts and may result in wounds of the pigs' tails, ears, flanks and legs. Such behaviours, which may lead to progressive tissue damage, are difficult to counteract.

On a commercial farm 22 groups of pigs with wounds on flanks (n = 16) and tails (n = 6) were matched with 22 control groups without wounds. All groups were provided with a novel rope, applied as a 'tail chew test'. Interaction with the rope was recorded semi-automatically about 45 and 120 minutes after introduction of the rope. Statistical analysis showed significant decrease of interest in the rope over time and significantly elevated interest in the ropes in pens containing wounded animals (median number of pulls per minute in control pens, flank-biting pens and tail-biting pens were 7.8^{a}, 10.2^{b} and 14.3^{b} respectively, where superscripts indicate significance levels ($P < 0.001$).

These results suggest that flank biting and tail biting increase exploration and destructibility in pigs. The approach taken is valuable in further understanding strategies to reduce injurious behaviours in pigs and improving pig welfare, e.g. by providing enrichment materials.

Introduction

Tail biting and tail docking generate major welfare concerns for pigs, especially those kept in intensive husbandry systems [1, 9]. Other injurious behaviours besides tail biting include flank, ear and leg biting.

In order to prevent tail biting most intensively-kept pigs are tail docked. Docking does not counteract flank- and leg biting.

Tail biting is regularly seen despite the current practice of tail docking. In the EU on average about 3% of docked pigs show tail lesions at the time of slaughter, but in undocked pigs as many as 6-10% may show tail lesions [9]. While tail biting has attracted considerable scientific attention (e.g. [18, 9]), much less is known about other injurious behaviours seen in growing pigs (ear-, leg- and flank biting).

Rope-based and sometimes (semi-)automated models of tail biting have been developed, mainly for the purpose of better understanding the causal mechanisms involved in a tail biting outbreak [10, 12, 13, 14, 15, 16]. Beattie et al. [2] used ropes to study known tail biting pigs, and Breuer et al. [8] investigated effects of breed and reported that gilts tended to manipulate a rope more often than boars.

Previously, we used a semi-automated novel object (rope) test to measure the pigs' interaction with the rope to measure the value of environmental enrichment [4, 5] and to measure the efficacy of tail-biting ointments [3].

While injurious behaviours are clearly multifactorial, inadequate enrichment appears to be a major risk factor [7, 9]. For example, providing long straw on the floor substantially reduced tail biting [20]. Relatively little, however, has been documented on the value of enrichment materials during outbreaks of other injurious behaviours such as ear and flank biting [20].

The objective of this study, therefore, was to examine the value of enrichment in pens with injurious behaviours such as tail, ear and flank biting by testing pigs' responses to a novel nylon rope in pens with and without wounded animals. A secondary objective was to evaluate the use of the semi-automated tail chew test as a tool to study pen-mate-directed behaviours.

Materials and methods

The study was conducted on a Dutch conventional farm rearing 1103 growing-fattening pigs housed in 14 highly similar units with pigs kept in mostly uniform and single-sex groups of 10 pigs per pen. All pens were 2.10 m deep and 2.30 m wide with a partly solid floor. All pens had a dry-feeder and nipples providing drinking mix (soluble food), a chain and a rope with a rubber flap (about 10x20 cm) hanging about 40 cm above the floor.

On the farm all units containing growing/fattening pigs were visually inspected from the feeding passage, searching for wounds on tails, flanks ears and legs, presumably due to injurious biting behaviours (hence excluding scratches due to fighting). When such a pen was found a control pen was selected without injured pigs. These pens formed matched-control pairs in which the tail chew test with a novel rope was done.

For the test a piece of white braided nylon rope (4mm in diameter; 48 cm long) was hung in all pens from the front wall, reaching up to about 30 cm above the floor of the pen.

Rope pulling behaviour was recorded semi-automatically (as described previously ([3, 4, 5], see also Figure 1) at two observation times, i.e. at about 45 and 120 min following introduction of the rope into the pen (T45 and T120 respectively). Two slightly different counter types were tested as part of a programme to improve the measuring technique (counters differed in the way they were protected from the pigs with or without a pvc container; the same counter type was used within pairs of pens).

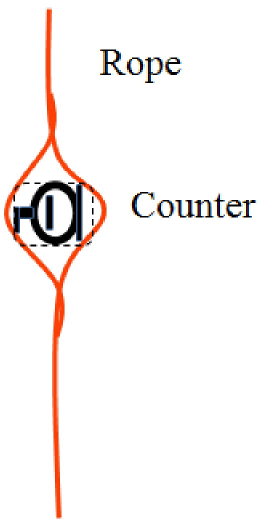


Figure 1. Schematic diagram of counter measuring rope pulling. The dotted line indicates 'logger type', i.e. extra protection of half of the loggers intended to better protect the logger from being destroyed by the pigs .

A mixed model analysis was performed using Genstat 11.1 [11] to determine the effects of observation time (T1, T2), wound type (flank biting, tail biting, control), gender (barrows; gilts, mixed sex), counter type (with and without pvc), unit, pair and their interactions on pulling frequencies (number of pulls per minute per pen). The response variable was analysed on the LOG-scale ($_{e}$log). Random effects for unit, case-control number/pair (within unit) and pen (within case-control) were included in the model.

The experiment was conducted in accordance with the Dutch and European legislation on the use of animals for scientific purposes. Since this was an observational study that did not negatively affect the welfare of the pigs no dispensation from the ethical care and use committee was needed.

Results

In total 22 pens were identified with wounds related to either tail biting (n = 6) or flank biting (n = 16). Out of 16 flank-biting pens, 4 pens contained only barrows and 12 contained only gilts. Out of the 6 pens with tail-biting wounds, 1 group was mixed-sex, 2 groups were all gilts and 3 groups were all barrows.

The analysis showed a significant effect of time (P = 0.01) and wounds (P < 0.001), and a trend for countertype (P = 0.09). Gender was not significant and neither did any of the interactions reach significance.

Pulling frequencies were higher at the first recording (T45) compared to T120 (predicted means on $_{e}$log scale were 2.47 and 2.21 respectively, sed 0.10). Control pens pulled significantly less frequent compared to pens with wounds due to flank biting and tail biting, while the two latter types of pen did not differ significantly (predicted means for controls[a], flank biting[b] and tail biting[b] (superscripts indicating significance levels) on $_{e}$log scale were 2.05, 2.32, 2.66 respectively, sed ranging from 0.097 to 0.19, with an average of 0.150). These values on $_{e}$log scale correspond with the values 7.8, 10.2 and 14.3 pulls per minute respectively.

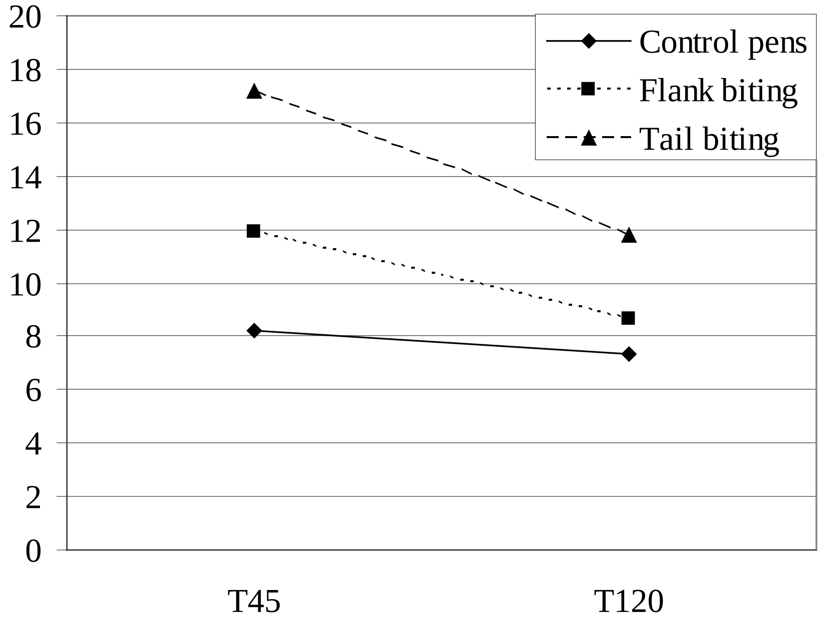


Figure 2. Backtransformed predicted means of rope-pulling frequency for control pens (n = 22), pens with flank-biting (n = 16) and pens with tail biting (n = 6), tested at two time points, 45 and 120 min. after introduction of a novel rope respectively (T45; T120).

Discussion and conclusion

Growing and fattening pigs in pens with biting wounds showed more interest in a novel rope as compared to matched controls.

Previously studies showed that pigs in barren pens had increased interest in novel objects [5, 19]. Hence, pigs in pens where injurious behaviours were evident may be experiencing their environment as more barren, hence be more interested in novel ropes and pen-mate-directed behaviour, leading to the injurious behaviour. Alternatively, however, injurious behaviours themselves are known to have a tendency to escalate due to reinforcing effects of the behaviour itself or its consequences (e.g. taste of blood). This may increase destructive behaviours generally, including rope manipulation (pulling hard). In other words, the present findings may relate to either a difference in cause or effect of the abnormal biting behaviour. In either case, the finding is in line with hypothesis that injurious behaviours are related to barren housing and may be solved by providing better enrichment materials [22], perhaps not only to prevent, but also to treat these behaviours as pigs in such pens seem to have a higher demand for enrichment materials.

The present results confirm the relationship between pen-mate-directed behaviours and pen-directed/exploratory behaviours (here tested as rope-directed behaviours; see e.g. [6, 17]), but as far as we know, this is the first quantitative studies reporting on flank biting as an injurious behaviour in growing-fattening pigs.

The finding that pigs from both tail and flank biting pens showed an increased interest in the rope, may indicate a similar etiology for both injurious behaviours i.e. increased interest in exploration.

The lack of a significant difference between flank biting and tail biting, despite the fact that predicted means for tail biting were almost as much elevated above the means for flank-biting pens, as these pens were elevated above control pens (see Figure 21) may be related to the fact that the number of flank-biting pens was much higher (16) compared to tail-biting pens (n = 6). This observation may also be related to the fact that the novel rope, which has previously been used as tail model as 'tail chew test' [3], morphologically resembles a conspecific's tail more than a flank, and hence may be expected to elicit a higher level of response in the case of tail biting compared to flank biting.

Previously, Breuer et al. [8] reported that gilts had a tendency to manipulate a rope more often than boars. Similarly, Zonderland et al. [21] found that gilts showed more tail biting. In this study we found no such effect for tail biting. In our small tail-biting sample (n=6) we found 3 groups of barrows, 2 groups of gilts and 1 mixed-sex group. For flank biting, however, the results seem to suggest a confirmation of the previous studies in that out of 16 pens with flank biting only 4 pens contained barrows and 12 pens contained gilts. Since there were considerably more pens with gilts compared to barrows on the farm (n = 57 versus n = 47), this effect was not significant (p = 0.087, Binomial test in Genstat 11.1).

Following our earlier work in relation to tail biting this study used a semi-automated rope model. It was confirmed that rope pulling behaviour may be a useful parameter to measure aspects of injurious behaviour in pigs. It may, therefore, not only be a suitable tail chew model, e.g. to test tail biting treatments [3], but it may also be useful to study related injurious behaviours such as flank biting.

Compared to earlier test applications (e.g. [3]) we here used a longer test duration (up to 120 minutes). This is related to on-going modifications directed at improving the test. In this study there was no substantial benefit of prolonged measurement as there was no specific effect of time (i.e. no interaction with wound presence), suggesting that there only was a general 'habituation' to the rope over time (but the pigs may also just get tired after a bout of exploratory activity). With respect to test optimalisation it was found that the modification of the counter type, that was directed at improved 'longevity', resulted, unfortunately, in a tendency for a lower pulling frequency in the better-protected recorders, perhaps indicating reduced sensitivity of the 'improved' design.

This study confirmed the suitability of the semi-automated rope test in helping to improve our understanding of injurious behaviours such as tail flank biting in pigs, and confirms that inadequate environmental enrichment may be implicated in injurious behavioural pathologies in pigs.

Acknowledgements

The authors are grateful for funding from the Dutch Ministry of Economic Affairs in support of the ANIWHA Era-Net initiative (FareWellDock project). We also greatly appreciate the hospitality of the farmer.

References

[1] Anonymous (2001a). Scientists' assessment of the impact of housing and management on animal welfare. *Journal of Applied Animal Welfare Science,* ***4****,* 3-52

[2] Beattie, V.E., Breuer, K., O'Connell, N.E., Sneddon, A., Mercer, J.T., Rance, K.A., Sutcliffe, M.E.M., Edwards, S.A. (2005). Factors identifying pigs predisposed to tail biting. *Animal Science,* ***80****,* 307-312.

[3] Bracke, M.B.M. (2009). Rope test may indicate efficacy of tail-biting ointments in growing pigs. *Animal Welfare, **18***, 263-266.

[4] Bracke, M.B.M. (2007). Multifactorial testing of enrichment criteria: pigs 'demanded' hygiene and destructibility more than sound. *Applied Animal Behaviour Science, **107**,* 208-232

[5] Bracke, M.B.M., Spoolder, H.A.M. (2008). Novel object test can detect marginal differences in environmental enrichment in pigs. *Applied Animal Behaviour Science, **109**,* 39-49.

[6] Bracke, M.B.M., Zonderland, J.J., Lenskens, P., Schouten, W.G.P., Vermeer, H., Spoolder, H.A.M., Hendriks, H.J.M., Hopster, H. (2006). Formalised review of environmental enrichment for pigs in relation to political decision making. *Applied Animal Behaviour Science, **98***, 165-182.

[7] Bracke M.B.M., Hulsegge, B., Keeling, L., Blokhuis, H.J. (2004a). Decision support system with semantic model to assess the risk of tail biting in pigs: 1. Modelling. *Applied Animal Behaviour Science, **87***, 31-44.

[8] Breuer, K, Sutcliffe, M.E.M., Mercer, J.T., Rance, K.A., Beattie, V.E., Sneddon, I.A., Edwards, S.A. (2003). The effect of breed on the expression of adverse social behaviour in pigs. *Applied Animal Behaviour Science, **84***, 59-74.

[9] EFSA (European Food Safety Authority) (2007). Scientific report on the risks associated with tail biting in pigs and possible means to reduce the need for tail docking considering the different housing and husbandry systems (Question No EFSA-Q-2006-029). *The EFSA Journal, **611***, 2-98.

[10] Fraser, D. (1987b). Attraction to blood as a factor in tail-biting by pigs. *Applied Animal Behaviour Science, **17***, 61-68.

[11] GenStat Committee (2000). *The Guide to GenStat*. Oxford, UK: VSN Int.

[12] Jankevicius, M.L., Widowski, T.M. (2003). Does balancing for color affect pigs' preference for different flavored tail-models? *Applied Animal Behaviour Science, **84***, 159-165.

[13] Jankevicius, M.L., Widowski, T.M. (2004). The effect of ACTH on pigs' attraction to salt or blood-flavored tail-models. *Applied Animal Behaviour Science, **87**,* 55-68.

[14] McIntyre, J., Edwards, S.A. (2002a). An investigation into the effect of different protein and energy intakes on model tail chewing behaviour of growing pigs. *Applied Animal Behaviour Science, **77**,* 93-104.

[15] McIntyre J., Edwards, S.A. (2002b). An investigation into the effect of tryptophan on tail chewing behaviour of growing pigs. *Proceedings of BSAS,* 34.

[16] McIntyre, J., Edwards, S.A. (2002c). Preference for blood and behavioural measurements of known tail biting pigs compared to control penmates. In P. Koene (Ed.), *Proceedings of the 36th International Congress of the ISAE*, p. 93. Egmond aan Zee, The Netherlands, August 7-10.

[17] Munsterhjelm, C., Peltoniemi, O.A.T., Heinonen, M., Hälli, O., Karhapää, M., Valros, A. (2009). Experience of moderate bedding affects behaviour of growing pigs. *Applied Animal Behaviour Science, **118***, 42-53.

[18] Schrøder-Petersen, D.L., Simonsen H.B. (2001). Tail biting in pigs. *Veterinary Journal, **162***, 196-210.

[19] Stolba, A., Woodgush, D.G.M. (1981). Verhaltensgliederung und Reaktion auf Neureize als ethologische Kriterien zur Beurteilung von Haltungsbedingungen bei Hausschweinen. *KTBL-Schrift, **264***, 110–128.

[20] Telkänranta, H., Bracke, M., Valros, A. (subm). Fresh wood reduces tail and ear biting and increases exploratory behaviour in finishing pigs. *Applied Animal Behaviour Science*.

[21] Zonderland, J.J., Wolthuis-Fillerup, M, van Reenen, C.G., Bracke, M.B.M., Kemp, B., den Hartog, L.A., Spoolder, H.A.M. (2008). Prevention and treatment of tail biting in weaned piglets. *Applied Animal Behaviour Science, **110***, 269-281.

[22] Zonderland, J.J., Bracke, M.B.M., den Hartog, L.A., Kemp, B., Spoolder, H.A.M. (2010). Gender effects on tail damage development in single- or mixed-sex groups of weaned piglets. *Livestock Science, **129***, 151-158.

Behavioural nutrition & big data: How geodata, register data & GPS, mobile positioning, Wi-Fi, Bluetooth & thermal cameras can contribute to the study of human food behaviour

B.E. Mikkelsen[1], A.K. Lyseen[1], M. Dobroczynski[2], and H.S. Hansen[1]

[1]Department of Development and Planning, University of Aalborg, Copenhagen, Denmark
bemi@plan.aau.dk
[2]sysCore ApS, Copenhagen, Denmark

Introduction

Place has traditionally been providing the conceptual and analytic platform for the studying physical environment and its relation to food behaviour [1, 2]. Where human activities previously were closely attached to the locations of home or work perspective of human activities has shifted towards a people-based view [3, 4]. Complexity and heterogeneity of human mobility no longer appear to correspond the use of residential neighbourhoods, but stress the need for methods and measures of individual activity and exposure, along with a change in awareness of researchers that human behaviour and activity is individual/people-based and not place-based [2, 4]. Portable intelligent devices has created a series of new opportunities for convenient assessment of food choice and as a result, food and lifestyle related behaviour is increasingly becoming the subject of measurement through application of such mobile devices [5, 6]. Intelligent devices such as smartphones, touch pads, etc. increasingly becomes used by consumers not only to get online using different wireless technologies but also for self-monitoring of lifestyle. For the research community this development also offers new opportunities since the devices can also be used in a reverse mode to track the behaviour of individuals applying GPS, mobile positioning, Wi-Fi and Bluetooth. Furthermore, thermal cameras offer new possibilities to track human behaviour, contrary to regular RGB-cameras where individuals are recognisable and subject for ethical issues. Application of GPS tracking within behavioural nutrition research [7] has been limited, but other research areas has embraced the technology and used it for task such as measuring travel patterns [8, 9], wildlife movement and habitats [10], exposure to toxics, pesticides or air pollution [11], following elderly with Alzheimer's and other dementia [12], and within health research to measure physical activity [13]. Use of mobile positioning, Wi-Fi and Bluetooth signals for tracking is more novel technologies in terms of tracking and less used in research for tracking than GPS. Application potentials of Bluetooth and Wi-Fi are comparable and has been used for measuring travel time through airport security [14], travel time on freeways [15] or mapping large crowds at mass events [16]. GSM has been sparsely used in for tracking with the exception of Ahas et al. [17], who used it for GSM tracking of tourists.

Use of the smartphones available in a study sample are believed to have large potential within social science, due to the use of apps and the large population penetration [18]. The penetration of smartphones worldwide has gone from 35% to 56% in two years and for cell phones in general to 91% [19] making the population basis for passive tracking with Wi-Fi, Bluetooth or mobile positioning profound. Normal RGB-cameras has long been used for surveillance but the ability to recognize people often poses a problem. Thermal cameras can replace RGB-cameras for tracking, as it is anonymized data. Traditionally nutritional research has used food diaries, questionnaires and interviews to analysis of food behaviour [20]. This is not uniquely to food studies, also transport research has previously frequently applied travel diaries, but implementation of GPS has reduced some of the shortcomings such as poor data quality, lack of reporting short trips, total trip times and destination locations [4, 8]. Functionalities related to Geographical Information Systems (GIS) offers simple representations of the physical environment including its opportunities for physical activity and food. The collection of behaviour information leads to the creation of very huge dataset, which induce problems for analysis. Hence, changing from place-based polygon features to people based point features requires different analysis and an increased focus on data administration and cleaning.

The contribution of this paper is on monitoring activity to improve the comprehension of food behaviour, which has numerous examples of placed-based studies [21, 22] and only a few people-based [7, 20]. This paper aims at giving an overview of the options available through these new technologies. The paper aims at an assessment of pros and cons for different type of tracking technologies and application setups. It gives examples of combinations of the technologies, and finally discusses the reach of these new opportunities along with a discussion of the ethical dimensions of such tracking.

Comparison of technologies – strengths and weaknesses

Deciding on which technology to use for a given research problem can be a straining problem, but very crucial to get right the first time. Knowing and understanding the strengths and weaknesses of the tracking technologies are the keys to selecting one or more technologies fitting the scale and environment of the research. The best fit of a technology in a study is influence by the environment, the extent of the study area, the required accuracy in the positioning, the need for respondents (active tracking) or for a only the movement (passive tracking) and the availability and pricing of the hardware and data. The environment for a study is often indoor, outdoor, or a combination. The GNSS and A-GPS technologies are only suitable for outdoor tracking. The thermal cameras require the environment to have open spaces with limited objects that block the view, whereas BT, Wi-Fi and mobile positioning copes with both outdoor and indoor tracking, regardless of the building layout and design. Only exception to this is, if the walls inside the buildings are blocking the short-wavelength radio waves and microwaves in for example BT and Wi-Fi, which is solved by adding additional BT sensors or Wi-Fi hotspots. The accuracy in estimated positions of the devices spans widely from approximately 1 metre to 20 kilometres dependent on the environment, hardware and signals. The accuracy of GNSS depends heavily on the environment of the study. With high buildings and narrow streets, the accuracy can easily be as bad as 20-30 meters dependent on the equipment, time of day and whether the equipment can use satellites from more than one system. In a bare field the accuracy can be as good as 1-5 meters again influenced by the same elements. The accuracy of BT and Wi-Fi can vary greatly based on two parameters, the range of the sensors and the possibility to triangulate between several sensors. The accuracy with no overlapping sensors will never be better than the scanning range of the sensors and due to the nature of the signals; the range of the sensors is a bit fuzzy making the exact precision of the positioning a bit uncertain. The accuracy of mobile positioning is the worst among the technologies mentioned in this paper. As mentioned in Ahas [17] there are several methods and network standards to base the tracking upon, of which A-GPS is the most accurate with 3 meters or better in open spaces and 20 meters in urban high-rise environments. Tracking solely on the cellular network would yield accuracies from under 100 meters to 20 kilometres, very much dependent on the environment and the density of the cellular network. Thermal cameras have proven to have a high position accuracy of 1-2 meters.

The size of the study area combined with the environment and the needed accuracy has a large impact on choosing the technology, the amount of devices needed and thereby the cost of the tracking. GNSS, A-GPS and mobile positioning are generally best suited for large areas as the individuals are carrying the technology necessary for tracking, while using BT, Wi-Fi and thermal cameras stationary sensors are needed. BT, Wi-Fi and thermal cameras are in theory plausible to use for large scale tracking at a very high price for equipment proportional to the amount needed or a low density of sensors only covering certain zones in the study area. Only covering all the supermarkets in an area could be an option if only the presence in a shop is relevant for the aim of the study. GNSS and A-GPS are the most accurate technologies to capture movement over a large area, but both requires respondents acceptance, while mobile positioning covers over 90% of the population worldwide without their permission is required. Choosing a passive or an active tracking technology both has consequences. Passive tracking technologies such as thermal cameras and mobile positioning have the potential for following close too every person. However, if BT and Wi-Fi are used for passive tracking, only a proportion of the population is registered, as only some of the customers will have either or both of the signals turned on. There are methods to bypass this and get a larger part of the customers tracked through inserting tags in shopping baskets and trolleys, encourage customers to turn on the signals with prizes or other benefits for the customer or combine BT and Wi-Fi to track on both signal types. If the goal of the study is to count the customers or measure the time spent at the shop, one or two sensors might be enough dependent on the amount of entrances and exits. With passive tracking only the movement patterns of individuals are the output, whereas with active tracking is it possible to join additional information about each person's health status, socioeconomic status, etc. However, active tracking requires the consent of each respondent. Willingness to participate in tracking varies across the population and the proportion that completes the study satisfactory are often in the range of 50-60%. GNNS and A-GPS are active tracking technologies only, whereas thermal cameras are unsuited for active tracking.

The technical knowledge needed to apply the technologies for tracking differ as GNSS have a huge amount of off-the-shelf solutions and delivers an output that needs little or no processing to be implemented in GIS. Mobile positioning, BT and Wi-Fi are all based on a cell structure, from which a processing is needed to change the data to point features with coordinates. A-GPS requires an app to handle the tracking, which needs development or buying and existing that fits the purpose and hardware. Thermal cameras require the most processing as movement detection is still far from commercial use. The technologies are in theory applicable worldwide, but in reality are GNSS, A-GPS and mobile positioning the only technologies, as the rest requires power supply or changing of batteries at regular basis. Mobile positioning is limited by the goodwill of the operating companies and the legal clauses in each country.nThe prize of a complete tracking setup is influenced by the amount of devices, the type of technology and the accuracy. In many cases, A-GPS probably is the cheapest option, as a large segment of populations already owns a smartphone and the technology is well suited for combination with BT and Wi-Fi tracking inside selected shops. Combination of the technologies is preferred, to utilize the strengths of each technology in different environment settings.

Discussion

Developments in wearable positioning technologies and GIS provide an opportunity for understanding and controlling many phenomena occurring in urban areas. The position technologies offer quantifiable measures of individual's movement and exposure as they make decisions in real-time. The degrees autonomy of people varies when making decisions about residence and work, who to socialize with and where to do that, as neither the individuals nor the living environment is static. It is unrealistic to assume that the majority of people spend all or most of their time in pre-defined geographical areas. The high levels of mobility in a population requires methods to measure this without limitations, as the residential or administrative boundaries, and the residential and working addresses not necessarily are the best identifier for our dietary behaviour. Residential and administrative boundaries may not provide the most adequate basis for analysis of the impact of place on health. Determination of the most appropriate scale for analysis of places influence on dietary behaviour or preferably, to apply flexible scales to fit the patterns of every person, would increase the understanding of individual and group behaviour. Potentially tracking technologies could help loosening the dependency of residential and administrative boundaries for representing place, by an individual measure of behaviour.

Traditionally dietary research has focused on questionnaires and interviews for analysis, but it is known that people have a tendency to embellish the reality about information on food intake. Likewise have self-reported information about trips and whereabouts been incomplete in especially short trips, start and end time of trips, and the addresses visited. The development of wearable positioning provides a more objective measure for the behaviour in terms of space-time information. The potential of the tracking technologies for research on dietary behaviour are great for measurement of exposure to example healthy and unhealthy food options. GIS provides the tools for combining the individual behaviour patterns with personal information on health and socioeconomic status and conduct statistical and spatial analysis. The purpose of implementing the tracking technologies in dietary research is not to replace the questionnaires and interviews on dietary behaviour, but perceived to add to a growing toolbox for researchers. Combining the data from questionnaires and interviews with tracking data increases the potential analysis potential.

Questionnaires, interviews and tracking are all invasive on the persons followed as the question presented by researchers and knowledge of being followed possibly affects the person's behaviour. Researchers ask for information that is often regarded personal or confidential as the whereabouts or food intake. This is active tracking and often requires a lot effort from the respondents, which also have a great influence on which segments of the population that are willing to participate. Passive tracking, on the contrary, removes the demands on the respondents and provides a measure of human movement and behaviour in a non-invasive manner.

Passive tracking often have the ability to capture the behaviour of large groups of the populations i.e. because of the high cell phone penetration in the population, but at the cost of losing the supplementary information of individuals. Both active and passive tracking offers great possibilities for research in term of tracking the general populations' movement patterns and shopping patterns or smaller samples of the populations' behaviour and its impact on health.

The choice of technology should not necessarily be limited to selecting one, as several of the technologies combined would enable tracking at several scales and accuracies to serve the objective of the study. Example, use of GNNS or A-GPS to track the overall movement of individuals in a city and then in all the supermarket and grocery shops set up Wi-Fi and BT tracking to track the movement inside the buildings. Then the tracking provides information of both the person's choice in food retailer and the choices made inside each individual food shop.

Use of wearable position technologies to follow the behaviour of individuals is an intrusion into people's privacy and violates many people's boundaries. Surveillance is in many countries frowned upon and raises a whole host of ethical issues for many people and governments. Privacy is a public concern, which causes debate about personal freedom and scientific ethics. Privacy, surveillance and data security are key aspects in both passive and active tracking. In active tracking, the person's identity is known to the researchers, who have the key responsibility, when storing, analysing and publishing results of tracking, to ensure the confidentiality of the participating persons. European and national legal regulations for data collection must be followed to ensure the persons and their locations are not identifiable. Privacy and ethical concerns influences the type of people willing to participate in studies involving tracking. Some of the concerns revolve around the fear of being subject to surveillance, and followed and listened to everywhere one goes. The growing ICT generation may be more open to the technologies, and the accustoming of positioning technologies may lower the concerns. Explanation of the aspects and demonstration of the results of tracking could decrease the concerns of privacy violations and data security. Private tracking have to treat the people in way that ensures their complete confidentiality also in the data. Thermal cameras ensure this compared to RGB-cameras, while application of other methods are needed for BT, Wi-Fi and mobile positioning. There are several methods [4] i.e. assigning 'dummy' variables to ID a device instead of using the MAC-address.

References

[1] Leal, C., B. Chaix. (2011). The Influence of Geographic Life Environments on Cardiometabolic Risk Factors: A Systematic Review, a Methodological Assessment and a Research Agenda. *Obesity Reviews,* ***12****(3)*, 217–230.

[2] Matthews, S.A., Tse-Chuan Y. (2013) Spatial Polygamy and Contextual Exposures (SPACEs) Promoting Activity Space Approaches in Research on Place And Health. *American Behavioral Scientist,* ***57****(8)*, 1057–1081.

[3] Miller, H. (2007). Place-Based versus People-Based Geographic Information Science." *Geography Compass,* ***1***, 503–535.

[4] Rainham, D., McDowell I., Krewski D., Sawada M.. (2010). Conceptualizing the Healthscape: Contributions of Time Geography, Location Technologies and Spatial Ecology to Place and Health Research. *Social Science & Medicine,* ***70****(5)*, 668–676.

[5] Ngo J., Engelen A., Molag M., Roesle J., García-Segovia P., Serra-Majem L. (2009). A review of the use of information and communication technologies for dietary assessment. *British Journal of Nutrition,* ***101***, 102-112.

[6] Thompson F.E., Subar A.F., Loria C.M., Reedy J.L., Baranowski T. (2010). Need for technological innovation in dietary assessment. *Journal of American Diet Associasion,* ***110***, 48-51.

[7] Zenk, S.N., Schulz A.J., Matthews S.A., Odoms-Young A., Wilbur J., Wegrzyn L., Gibbs K., Braunschweig C., Stokes C. (2011) Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study. *Health & Place,* ***17****(5)*, 1150–1161.

[8] Murakami, E., Wagner, D.P. (1999). Can Using Global Positioning System (GPS) Improve Trip Reporting? *Transportation Research Part C: Emerging Technologies,* ***7****(2–3)*, 149–165.

[9] Stopher, P., FitzGerald, C., Zhang, J. (2008). Search for a Global Positioning System Device to Measure Person Travel. *Transportation Research Part C: Emerging Technologies,* ***16****(3)*, 350–369.

[10] Markham, A.C., Altmann J. (2008). Remote Monitoring of Primates Using Automated GPS Technology in Open Habitats. *American Journal of Primatology,* ***70****(5)*, 495–499.

[11] Phillips, M.L., Hall, T.A., Esmen, N.A., Lynch, R., Johnson, D.L. (2001). Use of Global Positioning System Technology to Track Subject's Location during Environmental Exposure Sampling. *Journal of Exposure Analysis & Environmental Epidemiology*, ***11**(3)*, 207-215.

[12] McShane, R., Skelt L. (2009) GPS Tracking for People with Dementia. *Working With Older People* **13**: 34–37.

[13] Rodriguez, D.A., Brown, A.L., Troped, P.J. (2014) Portable Global Positioning Units to Complement Accelerometry-Based Physical Activity Monitors. *Medicine and Science in Sports and Exercise*, ***37***, 572–581.

[14] Bullock, D., Haseman, R., Wasson, J., Spitler, R. (2010) Automated Measurement of Wait Times at Airport Security. *Transportation Research Record: Journal of the Transportation Research Board*, ***2177***, 60–68.

[15] Haghani, A., Hamedi, M., Sadabadi, K., Young, S., Tarnoff, P. (2010) Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, ***2160***, 60–68.

[16] Versichele, M, Neutens, T., Delafontaine, M., Van de Weghe, N. (2012) The Use of Bluetooth for Analysing Spatiotemporal Dynamics of Human Movement at Mass Events: A Case Study of the Ghent Festivities. *Applied Geography*, ***32***, 208–20.

[17] Ahas, R, Aasa, A., Roose, A., Mark, Ü., Silm, S. (2008) Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study. *Tourism Management*, ***29***, 469–86.

[18] Raento, M., Oulasvirta, A., Eagle, N. (2009) Smartphones An Emerging Tool for Social Scientists. *Sociological Methods & Research*, ***37***, 426–454.

[19] Infographic. (2013) Mobile Growth Statistics | Digital Buzz Blog. <http://www.digitalbuzzblog.com/infographic-2013-mobile-growth-statistics/>. Accessed 28 February 2014.

[20] Kestens, Y., Lebel, A., Chaix, B., Clary, C., Daniel, M., Pampalon, R., Theriault, M., Subramanian, S.V.P. (2012) Association between Activity Space Exposure to Food Establishments and Individual Risk of Overweight. *PLoS ONE*, ***7**(8)*, e41418.

[21] Leung, C.W., Laraia, B.A., Kelly, M., Nickleach, C., Adler, N.E., Kushi, L.H., Yen, I.H. (2011) The Influence of Neighborhood Food Stores on Change in Young Girls' Body Mass Index. *American Journal of Preventive Medicine*, ***41**(1)*, 43–51.

[22] Liu, G.C., Wilson, J.S., Qi, R., Ying, J. (2007) Green Neighborhoods, Food Retail and Childhood Overweight: Differences by Population Density. *American Journal of Health Promotion*, ***21***, 317–325.

Making choices in space

Skov-Petersen, Hans[1] and Jacobsen, Jette Bredahl[2]

[1] Dept. of Geoscience and Natural Resources, University of Copenhagen
hsp@life.ku.dk
[2] Dept. of Food and Resource Economics, University of Copenhagen
jbj@ifro.ku.dk

Introduction

The term space is defined and applied in different ways though out academic disciplines and fields of application. In the present context we use the term environmental space in the sense of the physical spaces in which we navigate and move[1].

When humans move in environmental spaces, we make choices. For that purpose we need knowledge about the options afforded by the environmental space, in a given situation. The space can serve several purposes. Here we focus on 1) destinations where our needs can be satisfied, and 2) space as a movement infrastructure itself. Accordingly preference analysis of this movement has to take these two aspects into account. As noted by [18] (p119) who - as a point of departure - conceptualizes '… the spatial choice process as the subjective selection of the most preferred alternative from a subset of alternative,…', eventually including an error component in this choice. The present paper also uses the random utility theory to [8] suggest a framework for analysing spatial movement as a choice process. The framework combines movement based on 1) perceived information on the immediate surroundings and 2) on mental maps on one hand vs. a) movement restricted by an infrastructure (e.g. a transport network) and b) non-restricted movement (e.g. on an urban square or an open meadow) on the other. The paper presents observations and results related to real world applications of the framework in relation to bicycle and recreation behaviour.

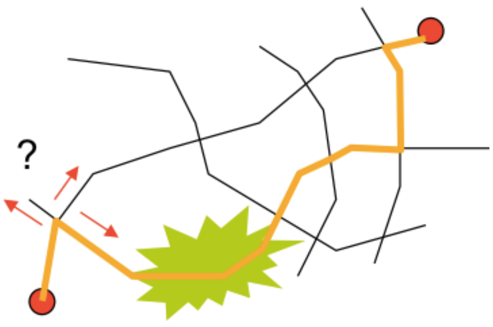
Perception, storage, cognition, adaption and application of spatial knowledge in relation to movement and navigation have been studied in a variety of scientific fields. Including psychology [10, 11, and 12], neuroscience /psychohysics [1, 9, and 13], behavioural geography [6 and 12], geoinformatics/GIS [2 and 17], and several fields of applied research including recreational behaviour [16], transport modelling [4] and econometrical choice modelling [19]. There is a tendency for the works within behavioural geography to acknowledge methods and theories from other scientific disciplines (mainly psychology and geoinformatics). One central consideration is the relation between the physical world and the subjective representation of perceived information from which spatial knowledge is gained and accordingly spatial decisions are made. Different people will apply the same information about their surroundings differently depending on for instance their personal characteristics (cultural, social, demographic etc.) and their motivation for moving. For a spatial context this cross-disciplinary paradigm is in particular addressed by the work carried out at University of California, Santa Barbara - spearheaded by the late Reginald Golledge (see e.g. [6]), but is also acknowledged in other fields, e.g. environmental economics [7]. The present work is regarded as part of the same strain of research.

As already indicated the concept of spatial movement decision-making and choices holds several connotations at different temporal and spatial scales. Spatial choices can be divided into choices of activity (e.g. to go the mall), mode (e.g. to walk), destination and route (see e.g. [3 and 5]). Emphasis of the paper will be put on individual human's navigation and path finding through different environments towards predestined destinations.

[1] We exclude notions of space aiming at metaphorical spaces (as in mathematics) and hyperspaces of the Internet [8]. Further – along with the distinction made by for instance [10 and 11] - we narrow down our scope to include only 'environmental' space and thereby exclude 'geographical' spaces (primarily representing the relative locations of regions and cities on Earth, 'figural' spaces (map representations) and 'vista' spaces (smaller location likes rooms town squares etc.).

We focus on what information can be elicited from revealed preference studies, where preferences are derived from subjects actual behavior. Consequently we can derive a preference for e.g. green space on a route from whether a respondent choose a route along a green environment more often than the alternatives

Montello [12] proposes that navigation takes place according to two spatial decision domains or scopes: Locomotion where navigation is based entirely on perception or our motor and vestibular systems. This is what you do when you are placed in an unknown territory with very little information about your surrounding except for what you get from your senses or corporal relocation. The other domain, Way finding is based on structured and comprehensive spatial knowledge in terms of a mental/cognitive map (which can be achieved from various sources including graphical maps, verbal descriptions, sensory stimuli (muscular or vestibular) combined by path integration [6]. The two approaches are illustrated in Figure 1.



1)

2)

Figure 1. Choice in a network: 1) navigation based on locomotion (perceived information) – choices are made between out-going edges at every node of the network. 2) navigation based on way finding (based on mental or screen/printed maps) – choices are made with regards to entire, alternative routes.

Analytical framework

RP analysis assesses the relation between observed behaviour and an identification of a set of alternatives (that were not applied by the subject). Accordingly, a main – and far from trivial – challenge is generation of such feasible alternatives. Feasible in the sense that they could actually serve as an alternative for the individual subject and also in that it is not possible to analyse an extremely large (depending on definition even infinite) amount of alternatives and consequently a decision has to be made on how to define possible networks.

Apart from the navigational domains (locomotion vs way finding), also the date model of the infrastructural domain has to be considered. Most research within transport choice modelling is based on motorized movement, and consequently movement is restricted by available roads. This may come short when less restricted transportation (e.g. pedestrians or animals) is considered. Here a continuous data field approach, based on cells, may be more appropriate. This is illustrated in Table 1.

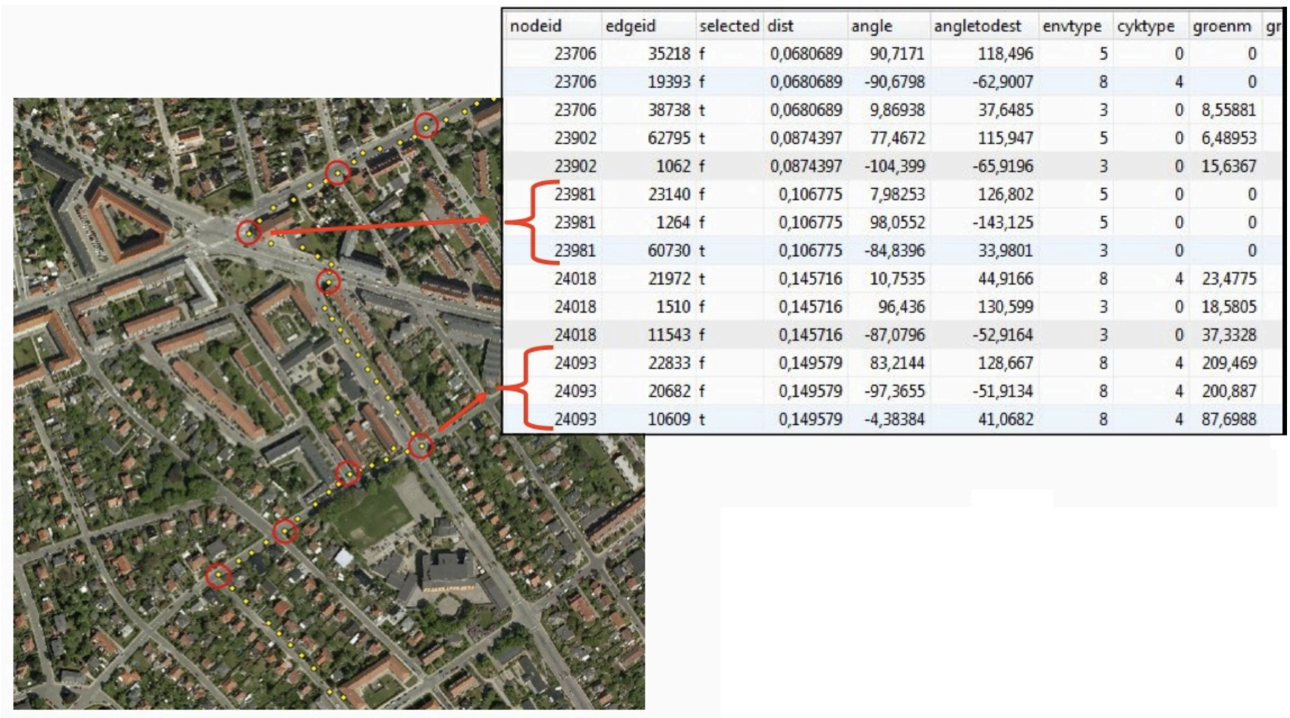
Table 1. A framework of analytical objects for assessments of spatial choices and decision-making.

		Infrastructure	
		a) Restricted	b) Free/open
Navigation domain	1) Locomotion	Outgoing edges from a node of a network	Cells representing next step in a raster
	2) Wayfinding	Routes (a series of edges) in a network	Routes (a series of cells) in a raster

The typical way of analysing choices in the random utility framework is by logistic regression, where the chosen route as well as the alternative(s) are described. Explaining variables might include spatial characteristics of the choice such as length, curvature, elevation profile, number of turns, land cover profile, but also socio-economic characteristics of the subject if available. Based on these choices the marginal rate of substitution between two characteristics can be calculated (resulting in willingness to travel or willingness to pay (or accept) estimates; elasticities can be calculated to predict behavioural changes and of course also the direct probability of a given choice over the alternatives can be calculated. . Such probabilities can be applied to e.g., agent-based simulation models to guide agents' spatial behaviour. In the following we describe in more details how this analysis can be carried out for the four possible situations described by infrastructure and navigation domain (cf Table 1), and in particular how the relevant choice set can be defined.

1a: Analysing outgoing edges from network nodes (locomotion/restricted)

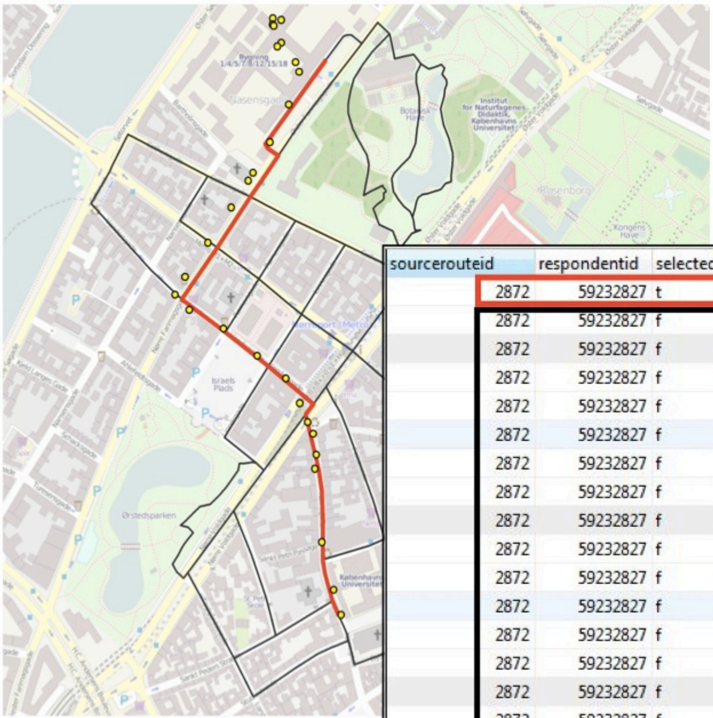
In this mode movement is modelled by choices made between edges spanning out from the choice location (i.e. the present node). In that case generation of alternatives to the selected edge is unproblematic. In addition to the attributes of the edges, a number of topological characteristics were assessed, - including the edges' relation to the destination and to the previous edge. Figure 2 shows an example of generation of such a choice set generated for bicyclists in Copenhagen [14]. The study included approximately 1800 trips taken by 179 bicyclists in Copenhagen. The results indicate significant effect of a number of the analysed variables, including both negative effects (for instance deviation from the bearing towards the destination and deviation from a direction 'straight on' regarding the previous edge) and positive effects (presence of bicycle facilities (tracks or lanes) and designated bicycle tracks). The approach is useful for identifying characteristics immediately observable from a point, but not for giving characteristics describing the entire route. Furthermore, interlinkages between nodes will have to be handled by e.g. lag-models.



nodeid	edgeid	selected	dist	angle	angletodest	envtype	cyktype	groenm	gr
23706	35218	f	0,0680689	90,7171	118,496	5	0	0	
23706	19393	f	0,0680689	-90,6798	-62,9007	8	4	0	
23706	38738	t	0,0680689	9,86938	37,6485	3	0	8,55881	
23902	62795	t	0,0874397	77,4672	115,947	5	0	6,48953	
23902	1062	f	0,0874397	-104,399	-65,9196	3	0	15,6367	
23981	23140	f	0,106775	7,98253	126,802	5	0	0	
23981	1264	f	0,106775	98,0552	-143,125	5	0	0	
23981	60730	t	0,106775	-84,8396	33,9801	3	0	0	
24018	21972	t	0,145716	10,7535	44,9166	8	4	23,4775	
24018	1510	f	0,145716	96,436	130,599	3	0	18,5805	
24018	11543	f	0,145716	-87,0796	-52,9164	3	0	37,3328	
24093	22833	f	0,149579	83,2144	128,667	8	4	209,469	
24093	20682	f	0,149579	-97,3655	-51,9134	8	4	200,887	
24093	10609	t	0,149579	-4,38384	41,0682	8	4	87,6988	

Figure 2. Example of a choice set representing bicyclists' choices based on locomotion. The Boolean attribute 'selected' (t/f) indicate if the option was chosen or not (i.e. for the first choice location indicated, edge 60730 was actually selected, whereas 23140 and 1264 where ignored. The U-turning edge was excluded from the choice set).

2a: Routes (a series of edges) in networks (way finding/restricted)
This approach is frequently applied in transport modelling. The route taken can be identified in terms of origin and destination and a set of characteristics (length, number of turns etc.). The main challenge is to generate potential alternatives to the selected route. A baseline route for comparison frequently applied is to use the shortest path. A more advanced – and potentially more realistic - approach is to generate a series of alternatives. A number of potential methods have been proposed. Several authors point at 'labelling algorithms' where the network is searched forward until feasible routes are found (see e.g. [15]). Others suggest the inclusion of a measure of the overlap between an alternative compared to the remainder routes of the choice set in terms of a measure of 'Path Size' [4]. Measures of overlap can also be used as a constraint to the search to ensure that routes in a choice set are spatial dissimilar [14], thereby allowing better estimation possibilities. Results – again based on 1800 trips taken by bicyclists in Copenhagen – include measures of the bicyclists' willingness to cycle additional distance to obtain or avoid given characteristics of a route. For instance, in [14], we find that bicyclists are willing to bicycle 20% longer if a bicycle track is available all the way than if it is not, and bicyclists are willing to drive 80 m extra to avoid a left turn on the route [14]. Compared to 1a), the advantage is the possibility of describing proportions of the route, and distance is taken into account explicitly.



sourcerouteid	respondentid	selected	length	nleftturns	nrightturns	nfrontturns	nbackturns	curviness
2872	59232827	t	1472,17	2	2	9	0	5,0266
2872	59232827	f	1395,66	4	4	6	0	8,605
2872	59232827	f	1363,76	3	3	6	0	6,6537
2872	59232827	f	1405,29	4	4	6	0	9,4421
2872	59232827	f	1458,33	3	4	8	0	8,0832
2872	59232827	f	1325,58	0	2	10	0	3,3621
2872	59232827	f	1405,46	3	3	8	0	7,1784
2872	59232827	f	1325,58	0	2	10	0	3,3621
2872	59232827	f	1470,97	2	2	10	0	4,8238
2872	59232827	f	1456,06	3	4	8	0	8,2176
2872	59232827	f	1410,24	4	4	6	0	9,527
2872	59232827	f	1385,97	3	4	6	0	8,6039
2872	59232827	f	1408,06	4	4	6	0	9,3798
2872	59232827	f	1445,02	1	3	10	0	5,3208
2872	59232827	f	1383,19	3	4	6	0	8,6656
2872	59232827	f	1325,58	0	2	10	0	3,3621
2872	59232827	f	1410,24	4	4	6	0	9,527
2872	59232827	f	1408,06	4	4	6	0	9,3798

Figure 3. Example of a choice set representing bicyclists' choices, based on way finding. The Boolean attribute 'selected' indicated if the option was chosen or not. The route displayed in red is the actual (map-matched) route (attributes are shown in the red box of the attribute table). Alternative routes are shown in black (attributes in the back box of the attribute table).

1b: Cells representing next step in a raster domain (locomotion/free)
Examples of such behaviour are hikers in nature, animals browsing for food etc. In the locomotion domain (basing decisions on perceived information) the 'next' step has to be identified directly as a subsequent point on the recorded track. Alternative locations can be identified in relation to the present point according to for instance distance and angle between the present and the subsequent points. An example of this approach applied to analysis of recreational behaviour will be given at the conference. The advantage of this method lies in its ability to describe slow movements, and movements where the destination is less important than the travel. The challenge lies in data explosion, as it becomes very large when cell size decreases.

2b: Routes (a series of cells) in a raster domain (way finding/free)

In a continuous field with no infrastructure, having a mental or representative (e.g. graphical) map of the taken route can be described as above. An example of such a movement can be a ship travelling in the open ocean or a hiker making his/her way off-piste to a predefined point in nature. To our knowledge no attempts until now has been reported in the scientific literature to create alternative routes for such an analysis. One option would be to perform a weighted random walk - including a preference weight for the bearing towards the destination - between the routes origin and destination applied by means of an agent base model.

Concluding remarks

Movement tracks can be applied to assess spatial movement behaviour, including preferences and route choices. The results, parameter estimates regarding characteristics of recorded routes, can be applied to a range of scientific questions: Can route preferences be assessed by RP analysis of tracking data? How do different types of agents find their way and decide on routes? What is the relative importance of different characteristics?

Strategically the direct results (i.e. the parameter estimates) can be applied to more informed mechanisms behind route-finding software (impedance setting) and further be applied to simulation of spatial behaviour (e.g. in relation to agent based modelling). In relation to planning, the results can be applied to assessment of the potential effect changes made to the infrastructure. However, for both of these purposes, it is important to be aware of the movement possibilities in terms of infrastructure, and the cognitive choices in terms of decision strategies. Especially the latter calls for more research in terms of testing different approaches against each other, but also by including knowledge from the psychological and (economic) choice set formation literature.

References

[1] Arleo, A., Rondi-Reig, L. (2010). Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. In Dolin, F.L., Mitchell, R.W. (Eds.), *Spatial cognition, spatial perception. Mapping the self and space*. Cambridge university Press.

[2] Batty, M., Crooks, A.T., See, L.M., Heppenstall, A.J. (2012). Perspectives on agent-based models and geographical systems. In Heppenstall, A.J. Crooks, A.T., See, L.M., Batty, M. (Eds.), *Agent-based models of geographical systems*, Chapter 1, p. 1-15. Dordrecht, The Netherlands: Springer Science+Business Media B.V.

[3] Bierlaire, M., Robin, T. (2009). Pedestrian choices. In Timmermans, H. (Ed.), *Pedestrian behaviour: Models, data collection and applications*. Bingley, WA, UK: Emerald group Publishing Limited.

[4] Frejinger, E., Bierlaire, M., Ben-Akiva, M. (2009). Sampling alternatives for route choice modeling. *Transport Research Part B: Methodological*, ***43****(10)*, 984-994.

[5] Golledge, R.G., Rushton, G. (1976). *Spatial choice and spatial behavior: Geographic essays on the analysis of preferences and perceptions*. Columbus, OH, USA: Ohio State University Press.

[6] Golledge, R.G. 1999. Human way finding and cognitive maps. In Golledge, R.G. (Ed.), *Human cognitive maps and wayfinding*. Baltimore, MD, USA: John Hopkins University Press.

[7] Lundhede, T.H., Jacobsen, J.B., Hanley, N., Strange, N., Thorsen, B.J. (2015). Incorporating outcome uncertainty and prior outcome beliefs in stated preferences. Forthcoming in *Land Economics*.

[8] McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P. (Ed.), *Frontiers in Econometrics*, p. 105-142. New York, NY, USA: Academic Press Inc.

[9] Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S.J, Frith, C.D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, ***97****(8)*, 4398-44-3.

[10] Meilinger, T. (2008). *Strategies of orientation in environmental spaces*. PhD Thesis. MPI Series in Biological Cybernetics, No 22. Max Planck Gesellschaft. Berlin, Germany: Logos verlag.

[11] Montello, D. (1993). Scale and multiple psychologies of space. In Frank, A.U., Campri, I. (Eds.), *Spatial information theory: A theoretical basics of GIS*, p. 312-321. Berlin, Germany: Springer-Verlag.

[12] Montello, D.R. (2005). Navigation. In Shah, P., Miyake, A. (Eds.), The Cambridge handbook of visuospatial thinking, p. 257-294. New York, NY, USA: Cambridge University Press.

[13] O'Keefe J, Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, OX, UK: Clarendon Press.

[14] Skov-Petersen, H., Snizek, B., Barkow, B., Harder H., Jacobsen, J.B. (Forthcoming). How do cyclists make their way? - A revealed preference study in Copenhagen.

[15] Skov-Petersen, H, Zachariasen, M., Kefaloukos, P.K. (2010). Have a nice trip: an algorithm for identifying excess routes under satisfaction constraints. *International Journal of Geographical Information Science, **24**(11)*, 1745-1758.

[16] Skov-Petersen, H., Gimblett, R. (2008). Introduction: Monitoring, simulation and management of visitor landscapes. In Gimblett, R., Skov-Petersen, H. (Eds.), *Monitoring, simulation and management of visitor landscapes*. Tucson, AZ, USA: The University of Arizona Press.

[17] Stanilov, K. (2012). Space in agent-based models. In Heppenstall, A.J. Crooks, A.T., See, L.M., Batty, M. (Eds.), *Agent-based models of geographical systems*, Chapter 13, p. 253-269. Dordrecht, The Netherlands: Springer Science+Business Media B.V.

[18] Rushton, G. (1976). Decomposition of space-preference functions. In Golledge, R., Rushton, G. (Eds.), *Spatial choice and spatial behaviour*. Columbus, OH, USA: Ohio State University Press.

[19] Wardman, M., (1988). A comparison of revealed preference and stated preference models of travel behaviour. *Journal of Transport Economics and Policy, **22**(1)*, 71-91.

Subjective versus objective captured Social Networks: Comparing standard self-report questionnaire data with observational RFID technology data

L. Thiele[1], M. Atzmueller[2], S. Kauffeld[1] and G. Stumme[2]

[1]Institute of Psychology, Department of Industrial/Organizational and Social Psychology, Technische Universität Braunschweig, Germany
{lisa.thiele,s.kauffeld}@tu-bs.de
[2]Knowledge and Data Engineering Group, University of Kassel, Germany
{atzmueller,stumme}@cs.uni-kassel.de

Abstract

Social network analysis provides insights into human social phenomena. Commonly, social network data is captured via surveys or newly via technical devices. In the described study, we combine subjective survey data with objective data from RFID tags in order to completely picture the closed network of an entire freshman students' cohort. RFID tags are small proximity sensing devices using radio frequency identification technology and can be worn as badges. Results indicate that the information of the two different data sources is similar but not entirely equal. Rather, they picture unique but complementing information and both should be considered in social network research.

Introduction

Social network analysis has become an important research perspective within various disciplines [4][14]. In particular, it enlarges the focus on individual attributes with the focus on interrelationships of people, which are present by nature as the individual is not living in isolation [6]. It offers the great possibility to study similarities, social relations, interactions, or flows between people [5], such as selection and influence processes [13]. For example, these processes are accountable for the formation of homogeneous subgroups in diverse teams, which may lead to subgroup polarization, increased conflicts and reduced performance [11]. The most common source of social network data is given by self reports [3]. However, some effort is done to examine social network patterns by means of observational data from sources such as emails [12], Bluetooth-enabled mobile phones [9], or Radio Frequency Identification (RFID) devices [1][3]. So far, the combination of both sources is used rather seldom, although this is promising for gaining a broader and more reliable insight into both, actually happened interaction and the emotional emphasis in this. Yet, Eagle et al. [9] combined and compared standard self-report survey data with observational data from Bluetooth mobile phones both regarding inter-individual proximity and found the information to be distinct and overlapping. Likewise, the goal of our study was to combine different sources of network data, in order to achieve an entire picture of social interaction including a comprehensive image of the occurred interaction and the underlying emotional weighting of the links between the examined actors. We followed a cohort of psychology freshman students during their first week at university and collected two types of network data (person-to-person interaction) using questionnaires and active Radio Frequency Identification (RFID) tags with proximity sensing. Hereafter, we will refer to data that was captured via surveys as 'subjective data' as opposed to behavioural data that is captured via RFID technology as 'objective data'.

Methods

Setting. The first week of freshman students at the examined degree program is usually organized as an introductory course before the regular courses. It was arranged over five days with a total attendance time of about twenty-five hours. The aim of the introductory week is to provide the newcomers with relevant information about the university, the degree program, and its contents as well as to introduce the professors and other lecturers, the departments/ chairs, and important committees. Moreover, this week as the first required course of the degree program offers a major opportunity to become acquainted with fellow students. In the year of data collection, three-fourths of the time (equals nineteen hours spread over five days) the week's events took place in a separate accommodation, which was suitable for the intended data collection and technically equipped for the purpose.

Sample. Seventy-eight students attended the introductory freshman's week (79.5% female). Sixty-seven of them belong to the new cohort of the psychology degree program. The remaining eleven students already studied two semesters at a different university and attended the week in order to get in easier. Those students were graded into the upper semester after the introductory course.

RFID technology. We employ RFID from the SocioPatterns consortium[1]. Worn as badges, RFID tags are able to identify face-to-face contacts by detecting other tags nearby. Criteria for contacts to be detected are (1) that the tags are facing one another (2) that they have a maximal distance of up to 1.5 meters, and (3) that this condition lasts for at least twenty seconds (see Figure 1).

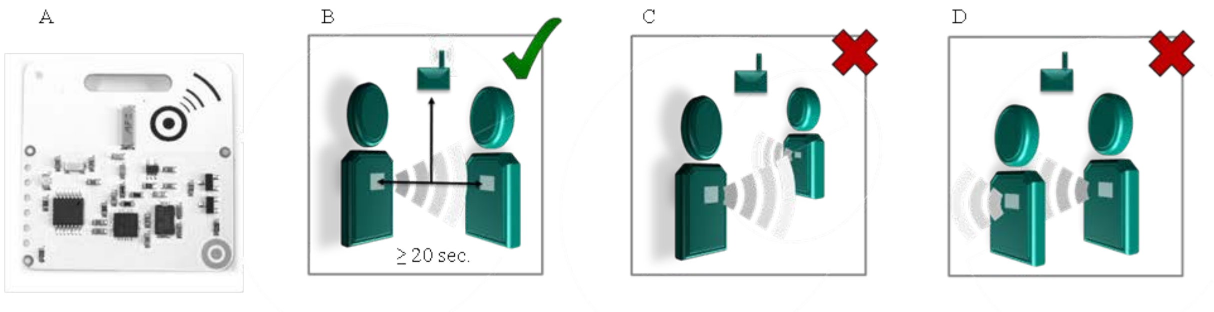


Figure 1. RFID-technology. (A) RFID-tag; (B) Accurate position of two RFID-tags to be detected as a contact; (C) Inaccurate position of two tags as they are not close enough to be detected as a contact; (D) inaccurate position of two tags as they are not facing one another.

When meeting these criteria, the tags send specific signals to RFID readers that are installed at fixed positions evenly spread all over the accommodation. The signals contain the unique ID of the sending tag, the ID of the detected tag as well as the time stamp of the contact between both tags. The readers then forward the signals to a server, where the information is stored and aggregated into a database, see Table 1 for a fictitious example. For detailed information about the functionality of RFID tags, see Barrat et al. [3]. With respect to the accuracy of the applied RFID tags, we refer to the results of Cattuto et al. [7] who confirm, that if the tags are worn on the chest, then very few false positive contacts are observed. In addition, according to Cattuto et al. [7] face-to-face proximity can be observed with a probability of over 99% using the interval of 20 seconds for a minimal contact duration.

Table 1. Data structure of RFID signals stored into a database with fictitious example data. SCR = sending RFID tag, DST = detected RFID tag, TSFROM = date and time the contact began, TSTILL= date and time the contact ended.

SCR	DST	TSFROM	TSTILL
1051	2987	21/10/2013 11:25:41	21/10/2013 11:26:51
1051	1652	21/10/2013 14:56:24	21/10/2013 15:01:04
1051	1652	22/10/2013 16:56:27	22/10/2013 16:58:57
[...]	[...]	[...]	[...]

Procedure. We asked each student to wear an active RFID tag while they were staying in the prepared accommodation. We combined the RFID tags with name badges to avoid multiple discomfort and to ensure unambiguous assignment during the five days of data collection. At the end of each day we requested the bagdes back so we could prepare them for the next day. In sum, we recieved 16780 data rows equal to the example in Table 1 via the RFID technology. These data include all detected contacts during the activities that happened

[1] http://www.sociopatterns.org/

within the accomodation, also including breaks and intervening periods. Moreover, the near surrounded areas such as smoking areas, the garden and the outer entrance area could be reached as well.

At the very end of the week, we captured the egocentric communication networks of the student among their fellows via questionnaires. We asked the students to select those fellow students on an exhaustive name list, with whom they interacted much during the introductory course. We transferred these data into a common network matrix with all names in the rows (respondent) and columns (persons to be selected), filled with zeros ('0'=not selected) and ones ('1'=selected).

Comparison of subjective and objective captured networks. We conducted two different analyses to compare the subjective and the objective captured networks. Both sorts of data were present for seventy-seven of the seventy-eight students. First, we examined the extent to which subjective and objective network data are matching. For this purpose, we converted the objective dataset into a format that is comparable with the subjective one. Therefore, we needed to reduce information. That is, we aggregated the contact duration and frequency over the whole week for each tag couple. We defined a contact to be meaningful (identical to the '1' in the subjective data matrix), when it was frequent above-average or durable above-average, respectively. The average duration of all contacts was 00:13:19 (hh:mm:ss), the average frequency of all contacts was 5.17 times. All contacts between pairs of tags beyond these cut-off values were coded '1', the rest was coded '0', separately for duration and frequency. Second, to get a more detailed insight into the differences between the diverse captured networks, we compared specific network measures using all existing information of the objective networks. For the subjective network we considered the directed network, which contains information about the direction of a tie. Furthermore, we constructed an undirected network, neglecting the tie information about who is sender and receiver, analogously to the objective network. On the one hand, we descriptively compared overall network measures such as the average degree of a node (number of direct links to others), the diameter (max. distance between two actors), and the density (ratio of the number of existing links to the number of possible links). On the other hand, we correlated the values of the individual network measures degree (number of direct links to others), betweenness (number of shortest paths from all actors to all others that pass through that actor), and eigenvector-centrality (centrality of an actor weighted by the centralities of linked others) of the diversely captured networks.

Results

We tested the matching of the subjective and the reduced objective network data by computing Cohen's Kappa [8]. Typically, Cohen's Kappa is used to assess the average agreement of two observers for example concerning their ratings of behavior [2]. In this case, we transfer this common statistic to our purposes as we treat the subjective and the objective data sources as two independent observers (3239 comparable ratings). Cohen's Kappa yielded a value of κ =.507 (subj. versus obj.$_{duration}$; 84.8% agreement) and κ=.485 (subj. versus obj.$_{frequency}$; 83.5% agreement), suggesting a fair congruence [10]. Moreover, we find that the overall network parameters differ according to their source (see Table 2 left).

Table 2. Overall network parameters of the different data sources. #Nodes = Network size (number of actors); #Edges = Number of links; Dia = Max. distance; (Avg.) Degree = (Average) number of direct links; Density = Ratio of #existing links to #possible links; Betweenness = paths passing through; Eigenv.-centrality=weighted centrality; Spearman’s rank order correlation coefficients: *p < .05. **p < .01, two-tailed.

	Overall network measures					Corr. of individual network measures		
	#Nodes	**#Edges**	**Dia**	**Avg. Degree**	**Density**	**Degree**	**Between -ness**	**Eigenv.-centrality**
objective	77	1622	3	42.13	0.55		objective	
subjective $_{directed}$	77	715	5	9.29	0.12	.281*	.333**	.262*
subjective $_{undirected}$	77	483	4	12.54	0.17	.267*	.265*	.272*

Naturally, the size of the networks (number of nodes) is the same. However, compared to the subjective network the number of connections, diameter, the average degree and the density of the objective network are rather different. We obtain a rather large deviation concerning the average degree of a node. Furthermore, the objective network is much denser than the subjective one. This indicates that the objective network covers more interactions during the observed time. On the individual level, we find significant but weak positive correlations of the subjective and the objective network measures (see Table 2 right). This indicates that the information of the two data sources is to a certain extent alike, but there is a substantial amount of uniqueness in the information of the diverse data sources.

Discussion

Results indicate that objective and subjective captured networks do not picture entirely the same, but can complement each other with respect to various aspects. Objective networks (RFID data) include all the contacts that took place in the equipped accommodation, which were not taken into account by the subjects in total when they filled the surveys. In contrast, the subjective networks (survey data) also contain interaction during the rest of the introductory week, covering places and times which could not be reached with the RFID technology. Simultaneously, the subjective data may be biased by several but relevant cognitive filters, for example by memory effects and emotional emphases. Our results suggest the complement use of both, self-reports and RFID technology. Advantages of this approach are formidable. First, it is possible to capture almost all interactions during the time of interest, no matter if the actors know each others names or even remember that they interacted with each other and no matter if the technical devices detected the interaction. Second, it is anyway possible to simoultaneoausly capture the subjective weighting of an interaction. That is for example, even if a pair of actors actually did not interact as much/long as another pair of actors, the interaction could be more meaningful to them, may resulting from the conversation's relevance or their interpersonal sympathy. Resultant, when it comes to capture the interaction network of large or newly composed groups and the researchers are interested in both a picture of actually happened interaction and the emotional emphasis of links between actors, both measures should be used. The use of each alone is to be considered in special situations of interest. When the group is very small, actors know each other and the emotional emphasis is of interest, subjective measures should be implemented. Contradictory, objective measures should be implemented when people do not know each other at all and the group is very large.

References

[1] Atzmueller, M., Benz, D., Doerfel, S., Hotho, A., Jäschke, R., Macek, B.E., Mitzlaff, F., Scholz, C. Stumme, G. (2011). Enhancing Social Interactionsat Conferences. *it+ti*, ***53(3)***, 101-107.

[2] Bakeman, R., Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge, UK: Cambridge University Press.

[3] Barrat, A., Cattuto, C., Colizza, V., Pinton, J.-F., V. den Broeck, W., Vespignani, A. (2008). High Resolution Dynamical Mapping of Social Interactions with Active RFID. *CoRR*, abs/0811.4170.

[4] Borgatti, S.P. (2003). *The state of organizational social network research today*. Department of Organization Studies, Boston University, manuscript.

[5] Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G. (2009). Network analysis in the social sciences. *Science*, ***323****(5916)*, 892-895.

[6] Brass, D.J. (2011). A social network perspective on industrial/organizational psychology. In Koslowski, S.W.J. (Ed.), *The Oxford Handbook of industrial and organizational psychology*. New York, NY, USA: Oxford University Press.

[7] Cattuto, C., den Broeck, W.V., Barrat, A., Colizza, V., Pinton, J.F., Vespignani, A. (2010). Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, ***5****(7)*, e11596.

[8] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, ***20***, 37-46.

[9] Eagle, N., Pentland, A., Lazer, D. (2008). Inferring Friendship Network Structure by using Mobile Phone Data. *PNAS*, ***106***, 15274-15278.

[10] Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd Ed.). New York, NY, USA: Wiley.

[11] Grow, A., Flache, A. (2011). How attitude certainty tempers the effects of faultlines in demographically diverse teams. *Computational and Mathematical Organization Theory, **17(2)***, 196–224.

[12] Kossinets, G., Watts, D.J. (2006). Empirical Analysis of an Evolving Social Network. *Science, **311***, 88-90.

[13] Veenstra, R., Dijkstra, J.K., Steglich, C., Van Zalk, M. (2013). Network-behavior dynamics. *Journal of Research on Adolescence, **23***, 399-412.

[14] Wasserman, S., Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, MA: Cambridge University Press.

The science behind measuring behavior

P. Pohl

Child Psychology Practice Garmisch, Garmisch-Partenkirchen, Germany
dr.pohl@kinderpsych-garmisch.de

A major development in 20th Century psychology and behavioral research was the adoption of the methods and statistical analyses of the English statistician, geneticist and evolutionary biologist R.A. Fisher (1890-1962) as the almost exclusive model of research practice and data analysis. The rapidity and thoroughness of this movement was remarkable as it comprehensively re-defined, within a period of only about 30 years (1925-1955), what „science" was, is, and ought to be in the field of psychological and behavioral research. Spreading from applied to experimental research, this paradigm gave the world factorial designs, randomization of research elements to treatment conditions, the analysis of variance (ANOVA), the analysis of covariance (ANCOVA), the null hypothesis, and null hypothesis tests of significance (NHST). Fisher's statistical tests were not the only practices adopted. His advocacy of factorial designs, i.e., experiments investigating more than one level of an independent variable, was equally influential, reflecting his view that experimental design and statistical analysis are only "different aspects of the same whole" [1]. Beginning in the mid-1950s and continuing up to the present, more than 90% of the articles published in psychological and behavioral research typically report the outcomes of significance tests [2], thereby demonstrating that the preoccupation with inferential statistics was and still is the main methodology in psychological and behavioral research with the underlying assumption that scientific knowledge emerges out of data through the application of statistical analysis.

Clearly, reproducible group data describe some kind of order and may well form the basis of a science. It cannot, however, be a science of individual behavior nor can it be a science of group behavior. It is, in fact, a science of averaged behavior of individuals linked together only by the averaging process itself [3] . Resulting from this state of affairs, disenchantment over the prevailing statistical methodology and its questionable relevance for understanding the behavior of the individual has led to a debate over the applications of statistics in psychological research, culminating in the establishment of a Task Force on Statistical Inference (TFSI) by the American Psychological Association (APA) in 1999. However, conspicuously lacking in the TFSI report is any recognition that single-case research designs might be an alternative to current statistical practices including statistical pattern recognition. This is curious because single-case research is based on a rejection of null hypothesis testing, emphasizes the visual analysis of data in graphs, and relies on replication as the key to making reliable causal inferences. Presumably the oversight is partly a result of the persisting confusion about the differences between case studies and single-case designs. Whereas case studies are usually clinical descriptions of indivdual cases and normally cannot support valid causal inferences, single-case research designs are controlled experiments from which valid causal inferences may be drawn. This neglect has deprived many researchers and practitioners in many areas of psychology and behavioral measurement of a powerful methodology for evaluating the effects of various psychological interventions and experiments. Among the numerous benefits of using single-case research methodology are that it helps researchers and prctitioners to focus on the individual (which is a widespread necessity in clinical and applied psychology), reduces some of the problems inherent in averaging across groups, makes it easier to undertake scientific investigations of rare and unusual cases and phenomena, and facilitates ethical innovation and professional accountability. At the very least, turning towards single-case methodology would help behavioral researchers to eliminate the „double standard" in their treatment of individuals, i.e. claiming to be able to apply science to individual organisms while working almost exclusively with group-aggregate data.

Why, then, has single-case methodology been used so rarely in behavioral research in the past decades? An analysis of 40 English-language introductory psychology textbooks published between 1979 and 2009 provides a clue: Investigators Hobbs and Chiesa summarized that „the history of psychology in the twentieth century as presented in introductory textbooks works to the detriment of behavior analysis and to the disadvantage of those students and professionals who would welcome and use behavior analysis if they were properly informed"[4]. In this context, it is important to mention that by about the middle of the 1950s, just as the so-called „cognitive revolution" began to gather momentum, behavior analysis and the single-case research methodology it had engendered, soon fell into disrepute and has been, for the most part, ignored by mainstream psychology to this day. It is no exaggeration to say that the issues surrounding this curious circumstance are fundamental to the science of behavior and will be dealt with in detail in the conference presentation.

And finally, based on the methodological argument developed here, human behavior studies employing *Observer XT* [5] and *FaceReader* [6] and animal studies employing *EthoVision XT* [7] and *PhenoTyper* [8] will be discussed with a view towards exemplifying the advantages of single-case methodology in measuring behavior.

References

[1] Fisher, R.A. (1935). *The design of experiments,* p. 3. London, UK: Oliver & Boyd.

[2] Hubbard, R., Parsa, R.A., & Luthy, M.R. (1997). The spread of statistical testing in psychology. *Theory & Psychology*, ***7***, 545-554.

[3] Sidman, M.(1960). *Tactics of scientific research*. New York: Basic Books.

[4] Hobbs, S., Chiesa, M. (2011). The myth of the cognitive revolution. *European Journal of Behaviour Analysis*, ***12***, 385-394.

[5] Reissland, N., Stephenson, T. (1999). Turn-taking in early vocal interaction: a comparison of premature and term infants′ vocal interaction with their mothers. *Child: Care, Health, and Development,* ***25(6),*** 447-456.

[6] He, W., Boesveldt, S., de Graaf, C., de Wijk, R.A. (2014). Dynamics of autonomic nervous system responses and facial expressions to odors. *Frontiers in Psychology,* ***5***, 1-8.

[7] Sprink, A.J., Tegelenbosch, R.A.J., Buma, M.O.S., Noldus, L.P.J.J. (2001). The EthoVision video tracking system – A tool for behavioral phenotyping of transgenic mice. *Physiology & Behavior*, ***73***, 731-744.

[8] Brielmaier, J., Matteson, P.G., Silverman, J.L., Senerth, J.M., Kelly, S., Genestine, M., Millonig, J.H., DiCicco-Bloom, E., Crawley, J.N. (2012). Autism-relevant social abnormalities and cognitive deficits in *Engrailed-2* knockout mice. *PLoS ONE,* ***7***, e40914.

Measuring the mating behaviors of free-ranging dusky dolphins (*Lagenorhynchus obscurus*)

D.N. Orbach[1], T. Kirchner[2] and B. Würsig[1]

[1] Department of Marine Biology, Texas A&M University at Galveston, Galveston, U.S.A.
dnorbach@gmail.com
[2] Institute for Biosciences- Marine Biology, University of Rostock, Rostock, Germany

Dolphins, whales, and porpoises generally have a polygamous (multi-partner) mating system [1]. Males use a variety of mating tactics to acquire paternity such as contest (e.g. fighting) and sperm competition [1]. Male mating strategies are conditional and depend on the distribution and monopolization potential of receptive females [2]. Among dolphin species, male mating tactics have been predicted based on sexual size dimorphism and relative testes mass in lieu of direct observations of mating events [1]. Mating behaviors are poorly described for most cetacean species and are generally limited to anecdotal reports with inadequate quantitative data [e.g. 3] or studies of captive animals with restricted inter-sexual interactions [e.g. 4]. It is logistically challenging to assess cetacean mating behaviors in nature as the animals are far off-shore and submerged underwater more often than they are visible to observers at the surface.

The dusky dolphin (*Lagenorhynchus obscurus*) population off Kaikoura, New Zealand (42° S, 173° E) provides a unique opportunity to assess cetacean mating behaviors and evaluate this population's mating system and tactics. There are ~ 2,000 individuals present at any given time, which form highly fission-fusion groups with continuous group size changes [5]. Foraging and mating behaviors are temporally and spatially separated [5-7]. Copulation events occur frequently near-shore during the breeding season in water visibility conditions that are good for above-water videography (>10 meters) [5, 8-10]. The mating system of dusky dolphins has been described as promiscuous with random breeding and polygyandrous (multiple males mate with multiple females) with non-random breeding [reviewed in 10]. We describe techniques to measure mating behaviors that will determine if free-ranging dusky dolphins demonstrate mate choice.

Dusky dolphin mating groups were followed from a 6 m research vessel during the peak breeding season from October through January in 2011-2012 and 2013-2014. Mating groups were defined as small isolated pods (< 50 individuals) with a constant group size in which either attempted copulations or an individual with an erect penis was observed [10]. Individuals in a group remained within 10 m of each other for an extended period of time [11]. Dusky dolphins in mating groups off Kaikoura typically perform simultaneous aerial leaps involving 1-3 individuals [10]. Mating groups were detected by three onboard researchers who scanned the horizon for leaping dolphins while travelling parallel to shore [9]. Follows were conducted in good boating conditions (Beaufort <3) where the dolphins could be tracked at ~ 5 m distance from the vessel. Follows ended when mating behaviors stopped, the group size changed, or the animals exhibited evasive behaviors towards the research vessel (e.g. resurfacing >50 m distance). Mating groups rarely evaded the research vessel. We adhered to the operating regulations set by the Marine Mammals Protection Act 1978, Marine Mammal Protection Regulation 1992, and local dolphin conservation guidelines [12]. No permit was required for this study.

Approximately 1,200 minutes of dusky dolphin mating video footage and 960 copulation events were video-recorded while travelling parallel to 91 mating groups. Continuous videos were recorded using a Sony Handycam HDR-XR550V recorder mounted to a chest-pod to reduce vibrations. Detailed *ad libitum* narrations included the group size, composition (adult, sub-adult, or calf), observed behaviors, initiation and termination of behaviors, and the sex of the dolphin performing the behavior. We calculated the number, duration, and frequency of copulation events and sex-specific mating behaviors. Males and females were distinguished based on observations of ano-genital slits and erect penises or ventrum orientation during attempted copulations [8-10]. Dusky dolphins copulate in ventrum-to-ventrum position with the female ventrum-down [5, 8-10]. We defined the duration of copulation bouts as the time between commencement and termination of ventrum-to-ventrum swimming. All approaches by the inverted male which resulted in ventral contact with the female were scored as copulations due to difficulties confirming ejaculation. An ethogram of sex-specific mating behaviors was developed based on published descriptions [10] and direct observations of mating groups (Table 1). We focused on tracking females in mating groups as their ventrum-down body positioning provided visibility of their individually distinctive dorsal fins [13]. It was challenging to detect subtle behavioral changes in males when they swam inverted beneath a female because of distortion of the water and increased distance from the researchers. When males were ventrum-down at the surface of the water, we could not confirm their sex. There were several males in a mating group and it was not possible to track all their behaviors simultaneously. Accordingly, our ethogram may be biased towards detecting female behaviors.

Videos were analyzed post-hoc using the video analysis software Transana [14]. Video playback speed was reduced to up to 0.1x the original speed. The behaviors in the videos were transcribed and time-stamped, including start and end times of dives and copulation events. The number of animals within one body length of the copulating pair was noted for each copulation. Preference was given to the *ad libitum* narrations when there was conflict with observable data on video. The annotations were summarized in1-minute time intervals to determine behavioral frequencies. Repeated occurrences of a single behavior were counted as separate events. The mean mating group follow duration was 13 minutes (S.E. ± 1.15, n= 91). Mating groups spent 92.6% of their time at the surface of the water, suggesting most of their mating behavior repertoire was observable. It is possible that different sex-specific mating behaviors occurred when the dolphins dove beyond our visibility range. Males and females copulated with several individuals in ventrum-to-ventrum contact. Males were inverted during all copulations and pushed the females against the surface of the water during 87% of all copulation events. Mating interactions generally consisted of five adult males chasing one adult female (S.E. ± 0.33, n= 90). Copulation events were brief, lasting 4.57 seconds in duration (S.E. ± 0.26, n= 90), and generally occurred once per minute (S.E. = ± 0.08, n= 91) per group until the group broke apart or mating behaviors ceased. On average, 1.3 presumed males were within one body length of the copulating pair (S.E. ± 0.09, n= 90). These males could cooperate or compete with the copulating male. Males performed several behaviors to maintain a close spatial position near the female such as swimming directly beneath (3% of male behaviors) or physically blocking (2% of male behaviors) the pathway of a copulating pair. Females exhibited maneuvers such as reorientation leaps out of the water (31% of female behaviors), rolling onto their dorsa so their ano-genital regions were inaccessible to males (11% of female behaviors), changing directions of travel (17% of female behaviors), and slapping their tails to terminate physical contact with a male (32% of female behaviors).

Our study provides direct assessment of the behavioral components of mating strategies for a free-ranging cetacean population. We contribute to the broader understanding of sexual selection theory and evolutionary mechanisms in animals with complex social and group dynamics. The observed male behaviors indicate exploitative scramble competition. During scramble competition, the most maneuverable or proximate male dolphins gain copulation opportunities with a female [9-10]. The males that copulate with the female closest to her ovulation remove her from the resource pool of rival males. The mating tactics of male dusky dolphins vary from other dolphin species. For example, some bottlenose dolphins (*Tursiops* sp.) exhibit mate-guarding and aggression [15]. We did not observe male dusky dolphins exhibiting physically aggressive behaviors, such as biting. Physically aggressive behaviors are associated with male contest competition [1], and when directed towards a female, can constrain her ability to escape or select alternative mates [15].

The observed female behaviors indicate an overall evasive tactic that potentially enables them to exercise mate choice. Females may extend the duration or difficulty of the mating chase to evaluate the competitive abilities of potential mates [10]. However, some females may not exercise mate choice. For example, females close to ovulation may instead actively reject all males and thereby induce intense male intra-sexual competition. The techniques to measure dolphin mating behaviors and the ethogram developed in this study set the stage for sequence analyses of female behavioral responses to male sexual behaviors, such as first and second order Markov chain analyses. Subsequent sequence analysis will enable us to differentiate between female mate choice and male rivalry and to determine if the mating system of dusky dolphins is consistent with a promiscuous or polygynandrous definition.

Table 1. Ethogram of sex-specific mating behaviors of dusky dolphins off Kaikoura, New Zealand.

Behavior	**Sex**	**Description**
Coordinated reorientation leap	Both	Simultaneous vertical leap out of the water by two or more dolphins followed by a head-first re-entry into the water. The whole dolphins' bodies clear the surface of the water
Reorientation leap of unknown sex	Unknown	A dolphin leaps vertically out of the water and re-enters head-first nearby. Its whole body clears the surface of the water. The sex of the dolphin is unknown
Tight circles	Both	Closely-spaced dolphins swim in a circle
Group dive	Both	All dolphins dive to depth and are no longer visible from the surface or an absence of surface behavior for at least 30 seconds duration
Inverted swim	Male	Male swims in a ventrum-up body position
Swim under leaping female	Male	Male swims inverted below a leaping female
Push female to surface	Male	Male pushes female up vertically while swimming ventrum-to-ventrum such that her dorsal region is above the surface of the water
Herd female against physical barrier	Male	Male pushes female against physical barrier (e.g. the keel of a vessel or the shore)
Swim under copulating pair	Male	Male swims inverted below and within one body-width of a copulating male
Interference	Male	Any movement by a male to directly break up a mating pair (e.g. positioning himself in the path of the copulating pair, rolling over the copulating pair, etc.)
Reorientation leap	Female	Female leaps vertically out of the water and re-enters head-first nearby. Her whole body clears the surface of the water
Tail slap	Female	Female raises her tail out of the water and strikes it against the surface of the water
Swerve	Female	Female quickly moves non-linearly through the water (e.g. small, fast changes in direction of travel)
Extreme directional change	Female	Female quickly changes her direction of travel 180° or more
Body roll	Female	Female rotates her body along her longitudinal axis (e.g. rolls onto her dorsum)
Caudal peduncle lift	Female	Female raises her caudal peduncle above the surface of the water during copulation
Spyhop	Female	Female positions herself vertically in the water with her head above the surface
Speed burst	Female	Female slices through the water at a high speed with minimal changes in direction

References

[1] Connor, R.C., Read, A.J., Wrangham, R. (2000). Male reproductive strategies and social bonds. In Mann, J., Connor, R.C., Tyack, P.L., Whitehead, H. (Eds.), *Cetacean societies: Field studies of dolphins and whales*, p. 247-269. Chicago, IL, USA: University of Chicago Press.

[2] Emlen, S.T., Oring, L.W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science*, ***197***, 215-223.

[3] Payne, R.S. (1995). *Among Whales*. New York, NY, USA: Simon & Schuster.

[4] Östman, J.S.O. (1985). *An ethogram for dolphin social behavior, and observations on changes in aggression and homosexual behavior among two subadult male bottlenose dolphins* (*Tursiops truncatus*) in a captive colony. M.A. thesis. San Francisco, USA: San Francisco State University.

[5] Markowitz, T.M. (2004). *Social organization of the New Zealand dusky dolphin*. Ph.D. thesis. College Station, USA: Texas A&M University.

[6] Würsig, B., Cipriano, F., Slooten, E., Constantine, R., Barr, K., Yin, S. (1997). Dusky dolphins (*Lagenorhynchus obscurus*) off New Zealand: status of present knowledge. *Reports of the International Whaling Commision*, ***47***, 715-722.

[7] Benoit-Bird, K.J., Würsig, B., Mcfadden, C.J. (2004). Dusky dolphin (*Lagenorhynchus obscurus*) foraging in two different habitats: active acoustic detection of dolphins and their prey. *Marine Mammal Science*, ***20***, 215-231.

[8] McOmber, L.C. (1999). *Mating activity among dusky dolphins (Lagenorhynchus obscurus) off Kaikoura, New Zealand*. B.Sc. thesis. Hanover, USA: Dartmouth College.

[9] Orbach, D.N., Packard, J.M., Würsig, B. (2014). Mating group size in dusky dolphins (*Lagenorhynchus obscurus*): costs and benefits of scramble competition. *Ethology*, ***120*,** 1-12.

[10] Markowitz, T.M., Markowitz, W.J., Morton, L.M. (2010). Mating habits of New Zealand dusky dolphins. In Würsig, B., Würsig, M. (Eds.), *The Dusky Dolphin: Master Acrobat off Different Shores,* p. 151-176. Amsterdam, The Netherlands: Elsevier Academic Press.

[11] Smolker, R.A., Richards, A.F., Connor, R.C., Pepper, J.W. (1992). Sex differences in patterns of association among Indian Ocean bottlenose dolphins. *Behaviour*, ***123*,** 38-69.

[12] Childerhouse, S., Baxter, A. (2010). Human interactions with dusky dolphins: A management perspective. In Würsig, B., Würsig, M. (Eds.), *The Dusky Dolphin: Master Acrobat off Different Shores,* p. 245-275. Amsterdam, The Netherlands: Elsevier Academic Press.

[13] Würsig, B., Jefferson, T.A. (1990). Methods of photo-identification for small cetaceans. *Reports of the International Whaling Commision*, ***12***, 43-52.

[14] Woods, D., Fassnacht, C. (2012). *Transana v.2.51*. The Board of Regents of the University of Wisconsin System, Madison, USA. <http://www.transana.org> [Last accessed on February 4, 2015].

[15] Connor, R.C., Smolker, R.A. (1996). 'Pop' goes the dolphin: a vocalization male bottlenose dolphins produce during consorthips. *Behaviour*, ***133***, 643-662.

Automated analysis of eye-body coordination during prey captures by newborn live-bearing fish.

Martin Lankheet, Twan Stoffers, Bart Pollux

University of Wageningen, Department of Animal Sciences, Wageningen, the Netherlands

Live bearing fish

Compared to egg-laying fish, live-bearing species invest much energy in the production of fewer, yet relatively more developed offspring. Although in many live-bearing species (e.g. mammals) viviparity is combined with extensive brood care after birth, this is not the case in the Poeciliidae, a family of freshwater fish that includes the well-known guppy (*Poecilia reticulata*), mollies (subgenus *Mollienesia*) and swordtails (genus *Xiphophorus*). The lack of post-natal parental care makes it critically important for newborn Poeciliids to quickly develop the required skills for survival, such as the ability to capture prey and escape predators.

In many of these behaviors visual perception and eye movements play a crucial role. To study the development of such behavior it is therefore critical to measure both swimming behavior and eye movements. In this paper we will show the development of prey capture behavior in newborn live-bearing fish, including changes in their use of eye movements for detecting and approaching prey. Since prey capture events are in many respects highly variable we recorded over 2000 capture events using high-speed video and developed fully automated analysis protocols for both body and eye movements.

Prey capture behavior

We studied ontogenetic changes in prey capture behavior during the first three days of life in the live-bearing fish *Girardinus metallicus* (Poeciliidae). Newborn fish need to start hunting immediately after being born in order to survive. It is not known, however, to what degree these behaviors are innately present at birth or acquired through learning during the first days of life. To measure these potential changes, and the underlying mechanisms, we have fed 28 newborn *G. metallicus* one baby brine shrimp (*Artemia nauplii*) at a time until satisfaction for a period of three days and recorded the strike and pre-strike behavior of every prey-capture event. During these three days the fish increased their food intake considerably (both in terms of the total amount of consumed prey and the rate of prey intake) and greatly reduced the rate of failures in catching prey. The question we address here is how the fish change their swimming pattern and eye movements while improving their hunting skills during this critical three-day period after birth.

Figure 1. Example frames at 4 ms apart from a prey capture event, just before ingestion of the prey (artemia).

High-speed movies

Prey capture events were filmed dorso-ventrally with a Mikrotron EoSens MC1362 high-speed camera at 500 frames per second and 1280x1024 spatial resolution. The fish were kept in petri dishes at 24 ºC temperature and a normal day-night light regime. LED arrays provided lighting from below (white LEDs) and from the sides (circular arrays of RGB LEDs), controlled independently by means of PWM signals generated by a MatLab controlled Arduino-based LED driver. Adjustment of different light sources allowed us to obtain the required contrast for body and fins relative to the background, and for the eyes relative to the body. Fig. 1 shows three examples of images from a recorded movie. In total we recorded and analyzed over 2000 capture events for 28 fish divided over the first three days of their lives. All experiments were approved by the ethical committee at Wageningen University (the Netherlands).

Data analysis

We used MatLab (including the image analysis toolbox) to analyze body, pectoral fins and eye orientations from the recorded frames. In addition, we tracked the position of the prey throughout the movie, to establish eye and head parameters relative to the prey.

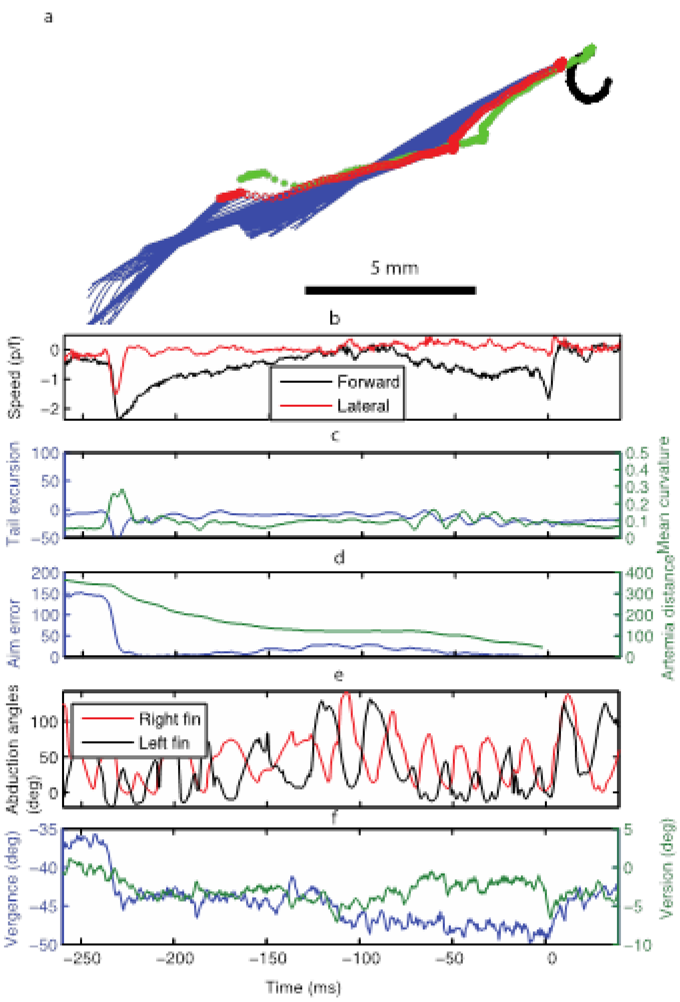


Figure 2. Quantifying prey capture events. a) body axes, midpoint of the eyes (red) and tip of the snout (green) in x-y coordinates, artemia in black. b-f behavioral parameters as a function of time relative to the strike (T=0); b) forward and lateral speed of the head. c) tail excursion and body mean curvature (arbitrary units). d) aim error of head relative to the artemia and distance to artemia. e) left and right fin abduction angles. f) eye vergence and version angles.

Previous studies on eye movements in, for example, larval zebra fish (*1*) have used the orientation of the main axis of the eyes to quantify viewing direction. For our images of live-bearing fish this was not optimal. In many movies the contrast between the eyes and the body was critically low and variations affected the length and or width measurement of the eyes. Such variations directly affected the estimate of viewing direction. To obtain a more robust measure we extracted the (mean) orientation of the flat, 'lateral' surface of the eye. First, the centroids of the eyes were determined by applying an adaptive threshold. Second, the center of the eye was used to define the outline of the fish that corresponded to the flat lateral surface of the eye. We determined the required outline of the fish by means of standard edge detection. Finally, eye orientations were quantified as the orientation of a straight line fitted to this part of the outline (i.e., orthogonal to viewing direction).

In addition to vergence (differences in orientations between the eyes) and version angles (mean viewing direction for the two eyes) we also determined body kinematics (axis curvature) and pectoral fin movements. The latter proved essential in explaining capture behavior, especially during the final strike. Pectoral fins were detected using edge-detection in a limited circle around the center-of-mass of the fish silhouette, followed by subtracting the silhouette. From the image of the fins we determined the fin abduction angles as the angle of the front-most edge relative to the midline through the head.

To compare measures for all recorded movies and between day 1, 2 and three we aligned all movies in time taking the last frame before the prey disappears into the mouth as the reference frame. This corresponds to the middle frame in Figure 1 and time T=0 in Figure 2.

Results

Figure 2 shows an example of analyzed data for one capture event. The maneuver starts with a C-bend and fast acceleration that aligns the head more or less with the prey (Fig 2d). Subsequently the fish slows down in the final approach to ingest the prey during a fast, forward acceleration, which is immediately followed by a strong deceleration. While approaching the prey the fish continuously increases its vergence angle, thus enlarging its binocular visual field. Version angles, on the other hand, are small and more or less constant during the approach.

Throughout the final approach the fish makes alternating left and right pectoral fin strokes. The changes in speed during the final strike are not accompanied by any obvious tail movements or changes in mean curvature. Instead, these speed changes strongly correlate with a change in the coordination of fin movements: Just before the strike the fish changes from alternating left-right strokes to a complete, simultaneous adduction (bringing both fins backward to align with the body). This symmetrical adduction causes a large forward thrust. The adduction is followed by a large amplitude, symmetrical abduction (moving fins outward and forward) that causes the nearly instantaneous break after the strike.

Based on a comparison of similar data obtained for day 1, 2 and 3 we conclude that the use of pectoral fins during the strike is highly stereotypical and invariant with age. The fish however, quickly learn to approach prey at higher speeds, while at the same time making less extreme eye vergence movements. Vergence angles show a similar rate of change, but maximum angles reduce significantly from day 1 to 2 and to 3.

Discussion

Fully automatic image analysis allowed us to quantitatively compare the performance and skills of newborn fish in prey capture events. It allows one to distinguish innate skills from skills that require development and calibration after birth. In zebra fish reflexes that control eye movements develop and become functional during the first days after fertilization (*2, 3*). Presumably, the gestation period in live bearing fish such as *Gerardinus metallica* is long enough to complete the circuitry for eye-body coordination. However, our data show that the final calibration to fine-tune eye-body coordination takes place after birth and presumably require visual input or practice. Our study further shows that other skills, such as coordinated movement of pectoral fins for final steering and for fast acceleration and deceleration during the strike, are already present and fully functional at birth.

References

[1] Bianco, I.H., Kampff, A.R., Engert, F. (2011). Prey capture behavior evoked by simple visual stimuli in larval Zebrafish. *Frontiers in System Neurosci* ence, ***5***, 101.

[2] Easter Jr, S.S, Nicola, G.N. (1996). The development of vision in the Zebrafish (Danio rerio). *Developmental Biology,* ***180***, 646-663.

[3] Easter Jr., S.S, Nicola, G.N. (1997). The Development of Eye Movements in the Zebrafish (Danio rerio). Developmental Psychobiology, ***31***, 267-276.

Measuring consciousness in turbot (*Scophthalmus maximus*) and common sole (*Solea solea*) subjected to electrical stunning before slaughter

M.B.M. Bracke[1*], A.H. Daskalova[3], J.W. van de Vis[2], E. Lambooij[1]

[1] Livestock Research, Wageningen UR, Lelystad, The Netherlands
[*] Marc.bracke@wur.nl
[2] Wageningen UR IMARES, IJmuiden, The Netherlands
[3] Faculty of Veterinary Medicine, Trakia University, Stara Zagora, Bulgaria

Keywords: Electrical stunning; Fish; Slaughter; Animal welfare; EEG; Behaviour

In order to limit painful experiences all animals including fish should be rendered unconscious before slaughter. We subjected two species of flatfish, turbot (*Scophthalmus maximus*) and common sole (*Solea solea*) to dry (de-watered) electrical stunning. In one group the fish were exposed to a short stun (1 s) followed by a second, long stun (20 s) after 1 min of recovery. In the second group the fish were exposed to a long stun (20 s). The electrical potential difference for the short stun was set to approximately 110-120 V_{rms} for turbot and 150 V_{rms} for sole. For the long stun the potensial difference was set to 50 V_{rms}. This resulted in a short-stun current of 2.3 ± 1.0 A_{rms} in turbot and 1.2 ± 0.7 A_{rms} in sole, while the long-stun currents were 4.0 ± 1.3 A_{rms} for 1 s + 1.0 ± 1.0 A_{rms} for 19 s in turbot in group 1, and 1.2 ± 0.6 A_{rms} for 1 s + 0.4 ± 0.2 A_{rms} for 19 s for sole in group 2. The objective was to examine whether the fish lost consciousness immediately (within 1 s) and did not recover when submersed in ice water immediately after the long stun.

In order to assess consciousness we recorded electrophysiological (EEG and ECG) and behavioural responses (pain sensation, breathing and response to vibration). EEG and ECG recordings were measured using a DI-720 data recording module with a WinDaq Waveform browser (Dataq Instruments, Akron, Ohio, USA; 250 Hz sample frequency). The EEG electrodes (20 mm long and 1.5 mm diameter; 55% silver, 21% copper and 24% zinc) were placed percutaneously. The ECG electrodes (the same composition) were placed subcutaneously, ventrally and dorsally of the upper pectoral fin. The earth electrode for both the EEG and ECG was placed subcutaneously near the tail. Pain sensation was recorded as behavioural and EEG responsiveness to the application of three scratches with a needle to the dorsolateral skin. For breathing gill movements were scored subjectively on a scale from 0 (no breathing) to 5 (large gill movements). Vibration was applied by three soft to moderate taps on the wall of the polystyrene box which contained the fish in the ice water. Pain sensation, breathing and vibration were scored at 12 time points between 0.5 and 75 min after the long stun. To our knowledge this was the first study in which responsiveness to vibration was compared to the behavioural and electrophysiological responsiveness to pain in electrically stunned fish.

Based on the EEG recordings all except one turbot (n=26) and one sole (n=10) lost consciousness after the short stun, and all but one turbot (n=22+13) and all but perhaps one sole (n=9+22) remained unconscious after the long stun. During recovery after the short stun behavioural responsiveness of the turbot followed EEG-based responsiveness. All sole, however, remained behaviourally quiescent after the short stun, while the EEG recordings indicated recovery in more than 60% of the fish. For at least 15 min after the long stun breathing (gill movements) was observed especially in turbot, while responses to vibration were especially seen in sole. Heart rates did not decline following stunning. Notably, after the long stun the fish tended to be more responsive to vibration as compared to the nociceptive stimulation.

These findings indicate that behavioural measures are not sufficiently reliable to determine unconsciousness during electrical stunning. Furthermore, our findings suggest that flatfish may remain responsive to vibration for longer as compared to nociception when submersed in ice water after electrical stunning, Finally, we propose that tail-first electrical dry stunning may be developed into a humane killing method for slaughter of turbot and sole.

A novel task to assess reversal learning in mice in a home-cage environment

E. Remmelink[1,2,3], A.B. Smit[2], M. Verhage[3], and M. Loos[1]

[1]Sylics (Synaptologics BV), Amsterdam
[2]Department of Molecular and Cellular Neurobiology, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam
[3]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research (CNCR), Neuroscience Campus Amsterdam, VU University, Amsterdam

Background

In several neurological and psychiatric disorders, executive functions, a collection of higher cognitive functions including attentional control, planning and flexibility, are affected. Measuring executive functions in mice is important in understanding the mechanisms underlying cognitive deficits in these disorders. The majority of currently available behavioral tests targeting these cognitive domains are operant tasks, which require extended training periods. Here, we describe a novel test for the cognitive measure ‘reversal learning’ in an automated home-cage environment called the PhenoTyper that circumvents extended training periods and the requirement of labor intensive animal-handling.

Methods

The task was implemented in an automated home-cage (PhenoTyper model 3000, Noldus Information Technology, Wageningen, The Netherlands), in which behavior was video tracked by a camera mounted in the top. Hardware actions were triggered by the position of the mouse (EthoVision HTP 2.1.2.0, based on EthoVision XT 4.1, Noldus Information Technology, Wageningen, The Netherlands). The cage was equipped with a shelter with two entrances, a feeding station, and spouts of a water bottle and a pellet dispenser. After an initial habituation period to the home-cage, a wall with three holes was placed in front of the pellet dispenser spout. For two days, mice had to learn to earn food (Dustless Precision Pellets, 14 mg, Bio-Serve, Frenchtown, NJ, USA) by going through the left hole in the wall (Initial learning). During this period the dispenser distributed 1 reward for every 5 times the mouse went through this correct left hole (FR5 schedule). The middle and right hole were the incorrect holes and passing these holes did not have any consequences. During the subsequent 2 days, the correct hole was switched to the right hole (Reversal learning). The number of passages needed to reach a criterion of 24 out of 30 passages through the correct hole, computed as a moving window, was used as a measure of performance during both stages. The number of passages through the left hole during the reversal stage provided a measure of perseverative errors. The number of passages through the middle hole represents a measure of random errors, because this hole was never rewarded [1]. The total number of passages, as well as total distance moved, were assessed as measures of general activity.
Experiments were carried out in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC), and with approval of the Animal Experiments Committee of the VU University.

Results and conclusions

C57BL/6J mice were able to attain the performance criterion of 80% correct within 1 day. Reversal learning requires suppression of the initially acquired response, while learning a new, competing, rule [2]. As expected, mice took significantly longer to attain the performance criterion during the reversal phase, compared to the initial learning phase. Nonetheless, all individuals were able to attain the criterion within the two available reversal learning days. Task activity was predominantly limited to the dark phase, suggesting that task performance had no impact on circadian rhythm. We have developed a 4-day protocol for reversal learning, which runs without human intervention. This task provides an easy and efficient way of screening mice for reversal learning in a home-cage environment, with improved animal welfare in comparison to conventional operant paradigms.

References

[1] Jentsch, J.D., Olausson, P., de la Garza, R., Taylor, J.R. (2002). Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology*, ***26***, 183–90.

[2] Dias, R., Robbins, T.W., Roberts, A.C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, ***380***, 69–72.

The difference is droll – Towards a cognitive bias test in mice

Lars Lewejohann [1], Vanessa Kloke [2]

[1]Behavioral Biology, University of Osnabrück, Osnabrück, Germany
[2]Department of Behavioural Biology, University of Münster, Münster, Germany

The quote from Oscar Wilde "The difference is droll /The optimist sees the doughnut /The pessimist, the hole" reflects the fact that that information processing is biased. The term 'cognitive bias' refers to such altered information processing resulting from an individual's background emotional state. Comparable to anxious or depressed humans, animals in a putatively negative emotional state are more likely to judge an ambiguous stimulus as if it predicts a negative event, while animals in a putatively positive emotional state are more likely to judge the same stimulus as if it predicts a positive event. By now, a number of different animal species have been demonstrated to judge ambiguous stimuli dependent upon their proposed emotional state. Thus, tests for cognitive bias have been suggested as a promising new indicator of animal emotion. Mice, including transgenic and knockout lines have become the most widely used species in biomedical research. To date, however, there are no reliable tasks for testing cognitive bias in mice that use reinforcers other than food and therefore would avoid confounding food motivation effects.

In the present study we aimed at establishing a cognitive bias test for mice using different approaches including spatial judgment as well as optical and tactile stimuli. We validated that our set-up can indeed assess different expectations about the outcome related to an ambiguous stimulus: mice having learned to expect something positive within a maze showed a more positive judgment of an unfamiliar location than animals having learned to expect something negative. In a second step, the use of spatial location as discriminatory stimulus was confirmed by showing that the mice's interpretation of an ambiguous stimulus depended on its spatial location, with a position exactly midway between a positive and a negative reference point provoking the highest level of ambiguity. Furthermore, in a pilot study we applied a spatial cognitive bias task to serotonin transporter (5-HTT) knockout mice, a well-established mouse model for the study of anxiety- and depression-related behavior. The anxiety- and depression-like phenotype of the 5-HTT knockout mouse model manifested - comparable to human conditions - also in a trend for a negative distorted interpretation of ambiguous information, albeit this effect was not statistical significant. The results suggest that the present cognitive bias test provides a useful basis to study the emotional state in mice, which may not only increase the translational value of animal models in the study of human affective disorders, but which is also a central objective of animal welfare research.

Ethical statement

The present work complies with current regulations covering animal experimentation in Germany and the EU (European Communities Council Directive 2010/63/EU). Experiments were announced to the local authorities and were approved by the Universities' ‘Animal Welfare Officers’.

Posters

Count your Chickens: Measuring Behaviour using a Series of JPEG Pictures

Erik Luc Decker[1] and Anja Brinch Riber

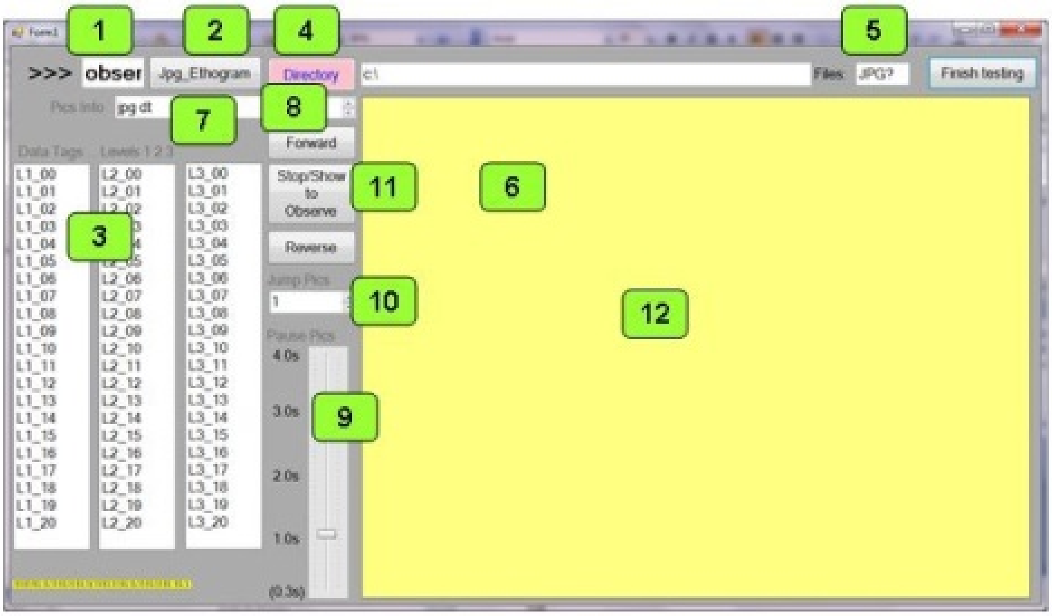
Aarhus University, Department of Animal Science, Tjele, Denmark
[1] erikl.decker@gmail.com

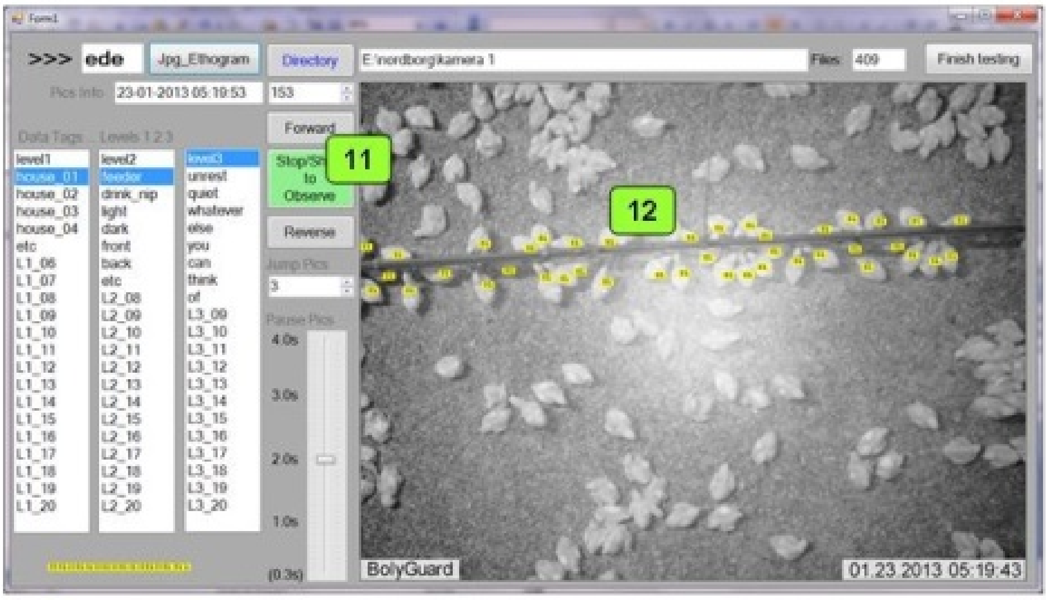
Background

Previous work has revealed that broilers use space heterogeneously, preferring the area along walls to the center of the house [1,2,3]. This results in high local densities, and may therefore increase severity of welfare problems related to high densities. Light is thought to have a great impact on the distribution of broilers [4]. Light emitting diodes (LED) are more efficient, durable, and retain the light intensity for considerable longer periods than the traditionally used fluorescent lighting (FL) [5,6]. We therefore hypothesized that by using LED lighting instead of fluorescent lighting, the distribution of light intensity would be more homogenous, and that this would instigate a more homogenous distribution of broiler, resulting in improved welfare and performance. The aim was therefore to investigate on farm the effect of LED lighting vs fluorescent lighting on spatial distribution in broilers.Thus, to determine the spatial distribution of broilers we wanted to count broilers in areas of known size (4.15-4.70 m^2) at different times during production runs. In addition, resource competition between the broilers was also investigated by counting the number of broilers drinking and feeding in these areas. Due to hygienic concerns the producer would not allow installation of video cameras in the broiler houses that would include up to 100 m long cables per camera. Game cameras powered by batteries were chosen instead. Typically they take pictures based on a motion detector, but the Bolyguard SG-560K camera also has time lapse, taking pictures every "x" minutes. This feature fitted well with our intentions of scanning every two hours. With a 16Gb SSD card and powerful batteries the cameras could be left in place for the six weeks of a production run.

Measuring behaviour

Data were collected from every second hour of the photoperiod on days 8, 13, 18, 23, 28, and 33 of age from six flocks of broilers housed in each of the two lighting environments. In order to alleviate the time consuming work of counting broilers manually on printed photos, a VB.Net program for counting on screen was written. In the program, when clicking on a broiler on a loaded picture a mark appeared. To make navigation easier in the large number of photo-files, some buttons (Forward, Reverse, etc.) were added, resulting in JPEG_Film.exe. This was further elaborated by reading the information of when the picture was taken, "DateTimeTaken", directly from the MetaData of the loaded JPEG file, resulting in the final program; JPEG_Film_Meta.exe. Using this program on a Windows8 tablet pc and with a stylus, generates the possibility of counting and logging the coordinates of an enormous number of broilers in a very short time. At the same time, tags can be added to the data, indicating location, behaviour, etc.





Using the program

Click and enter your initials [1]. Click and choose an ethogram file [2]. The 63 possible elements of the ethogram will appear here, divided over three levels. You can choose elements at any time. All clicks on elements will be registered in the dataset with the datetime stamp of the current picture [3]. Click and choose the directory with the pictures you want to use [4]. This shows the number of JPEG files in the chosen directory [5], loads the first JPEG picture [6], and shows its datetime stamp [7]. Click and enter the number of the JPEG you want to use and click [Stop/Show], or scroll up and/or down to it. While scrolling new pictures are not shown. Only when stopping the picture loads and its datetime appears [8]. Choose the length of the pauses between the pictures. The shortest possible pause depends on the pc used. If you choose longer pauses you can note events (or follow an animal) while the ”film” is running [9].

If you have a picture every 2 hours and you want to scan every 6 hours then jump three pictures [10]. Now run the "film" [Forward] or [Reverse]. click [Stop/Show] [11] on and count the observation you are interested in [12].

Results

The spatial distribution of broilers in the house did not differ between treatments, but it was found to be more homogenous the older the broilers were (treatment: $F_{1,390} = 0.82$; $P = 0.37$; age: $F_{5,391} = 43.95$; $P < 0.0001$). Likewise, competition over space allowance at the feed trough did not differ between treatments, but age affected the distribution of the broilers along the feed troughs (treatment: $F_{1,390} = 0.17$; $P = 0.68$; age: $F_{5,391} = 6.98$; $P < 0.0001$). The distribution of broilers relative to drinking nipples depended on both treatment ($F_{1,390} = 5.46$; $P = 0.02$) and age ($F_{5,390} = 15.35$; $P < 0.0001$); broilers in the LED treatment were more evenly distributed and the distribution was found in both treatments to be more homogenous the older the broilers were.

Try it!

The JPEG_film_meta program is open source, so if you want to try it contact erikl.decker@gmail.com.

References

[1] Arnould, C., Faure, J.M. (2004). Use of pen space and activity of broiler chickens reared at two different densities. *Applied Animal Behaviour Science*, ***87***, 153-170.

[2] Arnould, C., Fraysse, V., Mirabito, L. (2001). Use of pen space by broiler chickens reared in commercial conditions: ccess to feeders and drinkers. *British Poultry Science*, ***42***, S7-S20.

[3] Newberry, R.C., Hall, J.W. (1990). Use of pen space by broiler chickens - effects of age and pen size. *Applied Animal Behaviour Science*, ***25***, 125-136.

[4] Davis, N.J., Prescott, N.B., Savory, C.J., Wathes, C.M. (1999). Preferences of growing fowls for different light intensities in relation to age, strain and behaviour. *Animal Welfare*, ***8***, 193-203.

[5] Doe, U.S. (US Department of Energy). 2009. LED measurement series: LED luminaire reliability. <http://www2.unca.edu/environment/documents/luminaire-reliability.pdf>. Accessed 12 December 2013.

[6] Khan, N., Abas, N. (2011). Comparative study of energy saving light sources. *Renewable & Sustainable Energy Reviews*, ***15***, 296-309.

Transportation safety architecture: Measuring senses of safeness

M.D. Bredero

Saxion University of Applied Sciences
Kortestraat 27, 7419 CK Deventer, The Netherlands
mail@architectmaartendouwebredero.com

Introduction

Increasing mobility in cities requires additional and better stations for public transportation. This is motivated by the need to reduce congestion, but also because public transportation is much safer than travelling by car. The stations or so called 'hubs' are designed to accommodate various means of transportation. A better understanding of new modes of transfer is required before the actual design is carried out. Transferring from one mode of traffic to the other will be encouraged if the resultant streams of passengers can move freely and their sense of danger is reduced. This study focuses on bundling and streaming passengers from private into public space or more precisely, from a private into a public mode of transportation [1].

A new and specific model (transit hub for car/underground and bicycle/bus) allows research to be carried out on how to optimize the spatial and directional setting for providing unhampered and safe flow [2].

The hypothesis also states that optimal physical surroundings, combined with naturally lit outlet areas, compensates for the sense of danger or, better put, enhances the feeling of commuting comfortably [3].

Research questions

The questions to be considered are:

- How do passengers perceive bundling into transportation streams?
- Does a change of floor level in stations influence their sense of safeness?
- Does a longitudinal approach as opposed to a perpendicular approach induce different senses of safety?
- To which extent does architectural space with natural light enhance notions of safety?
- Does optimizing the situation require complete or partial separation of passenger flows and, therefore, zoning the stations in specific entities [4]?
- Does having an overview of the activities going on in and around a building in all cases improve the notion of safeness?
- To which degree does the station's orientation relative to the surrounding environment, including oncoming traffic, affect senses of safety?
- What is the relationship between human proximity and fear?
- To what extent can a virtual design model correlate with the actual physical building?

Type of research

The research involved is empirical with an inductive approach. Based on specific observations of behaviours, an attempt will be made, on the one hand, to deduce general rules. On the other hand, it is assumed, as a deductive hypothesis, that the presence of natural light will enhances the feeling of safety. Also, having an overview of the spatial surroundings generally seems to be appreciated. The research will consist of questionnaires being administered at specific intervals while respondents (arbitrary travelers and experienced commuters) are exploring the virtual environment. A survey will therefore be combined with a case study in order to test the hypothesis. For instance, the respondents have to answer questions on how safe and smooth (on a scale of one to five) they perceive the space they are virtually wandering through. Or how they appreciate having to change levels in order to get to the underground or their car. Recording their behavior during the simulated process of passing through and exploring the virtual architectural space will provide an insight into the psychological perception of spatial information. In this process, the respondents' physical condition, i.e. their eye movement, perspiration and heart rate are measured to register any corresponding level of anxiety. Depending on the number of measurements, the research will primarily be qualitative in nature. Only limited statistical calculations will therefore be necessary. The objective is naturally to achieve a neutral result by applying an appropriate methodology and precise focus.

Research methodology

Precise measurements can be achieved by working with two distinct groups, comprising: daily commuters using public transportation and novice users of public transportation. Within a given time frame, the subjects within each group are to explore the virtual models of the stations and respond to specific questions put to them. Set-ups A and B will then be examined and compared. The comparison will depend on the data gathered. Set-up A, the *horizontal* station: switching from car to underground and vice versa. Set-up B, the *split level* station: switching from bicycle to bus and vice versa. All subjects will first travel alone and then together in small groups during the test. It may be necessary to combine different types of participants, but this will depend on the results of the initial tests. Moreover, relevant questions will be put regarding comfort and orientation – i.e. the mental condition of subjects before and after the tests – with the answers being compared with their actual behavior [5].

Implementation of research

The current architectural Preliminary Design (see below) will be converted further into a semi-realistic environment using advanced visualization and gaming techniques. Subjects will explore the interior environment – either individually or in groups – by means of computer interfaces. They will interact with computer screens while sitting on chairs at gaming consoles. This can be done, for instance, at the advanced gaming facilities at the University of Twente (T-Xchange). This non-direct method (registration of mental decisions) will make it possible to measure the mental process of safety assessment.

General constraints

- Realistic model (architectural design model converted into real-life visualization)
- Representative measurement (large enough number and representative groups of subjects)
- Applicable (usable protocol for professional architects as a result)

References

[1] Doban, G., (2006) Designing for Safety. ENGINEERING.com LIBRARY, <http://www.engineering.com/library/articlesPage/tabid/85/articleType/ArticleView/articleID/85/Designing-For-Safety.aspx>. Accessed January 12, 2014.

[2] Several authors, (2012) CPTED Crime Prevention Through Environmental Design for Transit facilities

[3] <http://www.apta.com/resources/standards/Documents/APTA-SS-SIS-RP-007-10.pdf>. Accessed January 12th, 2014

[4] Walker, J. (2011). Human Transit. Island Press.

[5] Sartoretto, E., Boisvenue, T. (2010) Transit on two wheels: Building cycling into the public transit system.

[6] Journal of public transit in Ottawa. <http://www.transitottawa.ca/p/journal-of-public-transit-in-ottawa.html>. Accessed January 12th, 2014

[7] Green, J. (2013) Transportation Systems That Do More Than Move People. Dirt, American Society of Landschape Architects.

[8] <http://sustainablecitiescollective.com/dirt/176266/winners-urban-design-award-show-transportation-systems-can-do-more-move-people>. Accessed January 12th, 2014

[9] Transit Hub movie <http://ideemobiel.movinnio.eu/transit-hub/> via scan right.

Guiding principles for behavioural laboratory animal science:

A new resource for better science and animal welfare

P. Hawkins[1] on behalf of BAP, BNA, ESSWAP Foundation and LASA

[1]Research Animals Department, Royal Society for the Prevention of Cruelty to Animals (RSPCA), Southwater, U.K.
penny.hawkins@rspca.org.uk

Introduction

A new resource setting out *Guiding Principles for Behavioural Laboratory Animal Science* was published in November 2013, with input from over 50 experts representing a wide range of stakeholders: researchers, animal technologists and veterinarians from industry and academia, statisticians, laboratory animal breeders, scientific animal welfare organisations and regulators [1]. The project was supported and funded by the British Association for Psychopharmacology (BAP), British Neuroscience Association (BNA), the ESSWAP Foundation and the UK Laboratory Animal Science Association (LASA), all of which were represented on the Steering Panel.

The Guiding Principles are designed to help make informed decisions about the best way to carry out studies of animal behaviour in biomedical experiments, with respect to ensuring that work is justified, implementing the Three Rs (replacement, reduction and refinement) and improving scientific validity. They are intended to be useful to anyone with an interest in behavioural laboratory animal science, including researchers (as both an introduction to those new to the field, and as a refresher to those who have already gained experience) and members of ethical or animal care and use committees.

Content of the Guiding Principles

The Steering Panel recognised that it would be impossible to produce detailed recommendations for individual procedures, because so many different types of procedure are undertaken, in varying experimental contexts, to study a wide range of species at different stages of development. Instead, the document sets out generic Principles, most of which apply to all species, in some cases dealing specifically with mice and rats as these are particularly commonly used. The main focus is on laboratory research, although many recommendations also apply to ethological research in the field, and the recommendations are intended to have worldwide relevance, rather than focusing on any particular country or legislation.

The Guiding Principles are divided into seven sections, which highlight important questions that should be resolved before, during and after studying the behaviour of laboratory animals.

The 3Rs and Ethical Evaluation advises on factors to consider at the project design stage, to ensure that the work will meet high scientific and animal welfare standards. It defines the 3Rs, includes guidance on assessing both harms and benefits, and addresses record-keeping and (of special relevance to those working within the Europan Union, EU) retrospective assessment of projects and assessment of actual harms.

Justifying studies of Laboratory Animal Behaviour discusses how to assess scientific validity, distinguishing between translational, predictive, construct and face validity. It also outlines ways of dealing with complicating factors such as co-morbidity.

Choosing the Procedure incorporates a list of questions that aim to prompt a critical appraisal of the proposed work, including practical factors and a focus on each of the 3Rs.

Training provides guidance on ensuring that the experimental work is carried out by competent investigators. Basic legal requirements for training and competence (e.g. as set out in the EU Directive 2010/63/EU) address generic skills, but evaluation of animal behaviour often requires more specialist skills and the document offers suggestions on how to ensure competence in these.

The Animal explains how differences in animal behaviour due to species (including whether inbred or outbred), strain, age, sex and the source of the animal can lead to variation that could undermine validity. This section gives examples and discusses how to consider these when choosing the most appropriate animal to achieve the objectives.

The Environment gives examples of environmental factors that can affect animal behaviour. These include housing, stocking density, cage environment (e.g. position in the rack, enrichment) and the facility environment (e.g. noise levels, light). It discusses whether and how control for environmental factors, taking both animal welfare and science into account.

The Experiment & Analysis of the Data re-emphasises the need to plan the statistical analysis of the data and the experimental design simultaneously – i.e. before starting the experiment. It also suggests ways of avoiding subjective or systematic bias when gathering or interpreting the data and warns the reader not to make unjustified assumptions about causes and consequences of changes in animal behaviour.

How to Obtain a Copy

Guiding Principles for Laboratory Animal Science is free to download from the ESSWAP website (www.esswap.org) at http://www.esswap.org/guiding-principles-for-behavioural-laboratory-animal-science.html and is also available from the BAP, BNA and LASA websites.

The Steering Panel is especially keen to receive feedback about the *Guiding Principles* and also intends that the document should be reviewed and revised periodically – this version is regarded merely as a first edition. The Panel would welcome your comments and suggestions at info@lasa.co.uk (please put 'Guiding Principles for Behavioural LAS' in the subject line).

Reference

[1] BAP, BNA, ESSWAP, LASA (2013). *Guiding Principles for Behavioural Laboratory Animal Science (1st Ed.)*. Available at http://www.esswap.org/guiding-principles-for-behavioural-laboratory-animal-science.html. Last accessed on March 6, 2014.

Results of the Open Field Test at different light intensities in C57 mice

F.J. Martin-Arenas[1] and C.O. Pintado

Centro de Producción y Experimentación Animal/CITIUS, University of Seville, Seville, Spain
[1] animalario@us.es

Introduction

There is widespread agreement on the need for standardization of parameters that might affect behavioral data generated in phenotyping experiments in different laboratories in order to minimize the variability of results and waste of animals. The Open Field Test, as well as other tests for measuring anxiety and locomotion behaviors in mice, is difficult to standardize between laboratories and the results obtained are often contradictory. However, this should not prevent us from making an effort to control and delimitate conditions that could potentially increase the variability obtained from these tests in order to reduce the use of animals. Even though the level of illumination is generally considered to be an important factor that may influence the mood of animals when confronted with a novel space, particularly in nocturnal animals such as mice, there is a considerable lack of information on light intensities used in such studies throughout the bibliography. In the cases where information is provided, variation of light levels used is surprisingly broad (Table 1).

Table 1. Review of 100 recent articles in relevant journals using Open Field or Plus Maze tests.

Most of the articles do not mention the light intensity that they are using to carry out the experiments.	67
Light intensity indicated (but ranged between 2 and 657 lux).	19
The light intensity is not indicated but the setup used to test mice is described (which varied widely: 1, 2 or 3 bulbs, power, color and different distances from the arena).	14

We tried to address whether and to what extent light intensity influenced the anxiety and exploratory related behaviors of C57 mice in our Open Field Test in control conditions and after anxiolytic or anxiogenic treatment.

Materials and methods

Experiments were conducted in accordance with approved IACUC protocol (CEEA- University of Seville) and performed following the welfare standards of the European Union.

Light intensities chosen: 40 lux, approximately the light that reaches the interior of the animal cages measured with a calibrated luxometer. 250 lux, approximately the light intensity in the animal room. The animals are usually only exposed to this light when transferring from cages or during manipulation. 600 lux, animals are not exposed to this light. In some experiments is used intentionally as an aversive stimulus [4], but some studies are performed under this and even higher levels of illumination [5].

Animals: 180 Adult (9-10 weeks old) male and female C57BL/6 mice bred in our facility were used; they were housed in a room with a 12-hour light/dark cycle (7:00-19:00). Room temperature was fixed at 21 ± 2ºC. Food and water were available ad libitum in all cases. Animals were randomly distributed in 9 groups (Control-Epinephrine-Diazepam, at the 3different light intensities chosen).

Open field test: The open field apparatus used in this study measured 50x50x50 cm., made of grey plexiglass, and was illuminated by 4 bulbs of 60 W placed at 1,8m above the arena; a potentiometer allowed us to adjust the intensity of light reaching the bottom of the maze as desired.

Variables studied: Selected parameters were collected during the test. N entries (frequency of mouse entry in the arena center), Ptime periphery (permanence time in the periphery), Dist periphery (distance travelled in the peripheral zone), Total dist travelled (total distance travelled in the test) and Resting time. Ethological variables were also observed and taken into consideration.

Drugs: Epinephrine (EP) 1mg/Kg as anxiogenic treatment [6], or Diazepam (DZ) 5mg/Kg as anxiolytic treatment were injected intraperitoneally 30 min before the behavioral test [4]. The same volume of saline was injected into the control groups.

Procedure: Prior to the experiment mice were taken to the test room to acclimate them to the test room conditions. Mice were treated with EP, DZ or saline 30 min before the test and were placed anywhere in the open field. The arena was divided in two zones, the peripheral zone next to the open field's walls and the central area which were similar for all groups tested. Behaviors were recorded for 30 min using a video camera and video tracker. Results were exported to the statistical package SPSS for calculations. The open field was carefully cleaned after every test with a non-odor detergent.

Statistics: Data from the behavioral assays were subjected to analyses of variance, ANOVA, Tukey-Dunnet test (homogeneus variances or not), Kruskall-Wallis followed by Mann-Whithney U-test when distributions were no normal.

Results and discussion

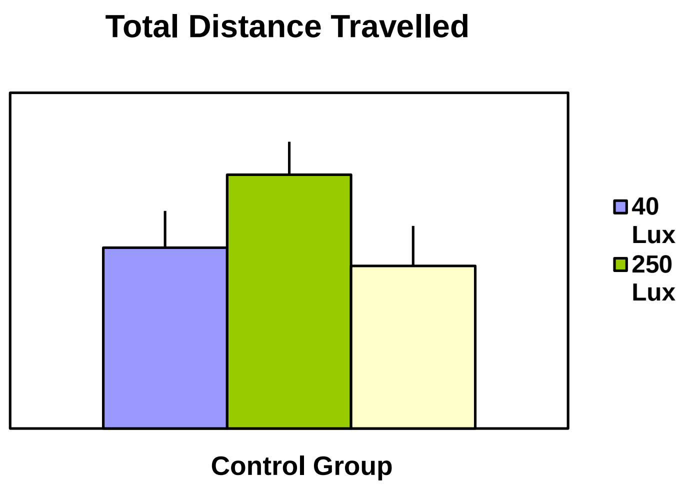
As shown in table 2, highly significant differences($p<0.01$) were observed when the maze was brightly illuminated (600 Lux) compared to 250 lux and 40 lux both in the number of entries and in the relative time spent in the periphery of the arena in both sexes. Interestingly, in females distance travelled was statistically different between 600 and 250 Lux, but not when comparing 600 and 40 lux (Table 3); this seemed to be related (figure 1 and 2), to the exploration becoming similarly reduced (and resting time increased) when animals are either relatively calm or subjected to aversive stimuli although due to opposite mood states (resting versus freezing) and could mean that females are more susceptible to the increase in the intensity of light.

Table 2. Effects of light intensity within control and treatment groups. ▲▼($p<0.01$), ▽△($p<0.05$), - (Not significant).

MALES	CONTROL			EPINEPHRINE			DIAZEPAM		
	40-250	250-600	40-600	40-250	250-600	40-600	40-250	250-600	40-600
N entries	-	▼	▼	-	-	-	-	-	-
P time periphery	-	▲	▲	-	△	-	-	-	-
Dist periphery	-	-	-	-	-	-	-	-	-
Total Dist travelled	-	-	▼	-	-	-	-	-	-
Resting time	△	-	▲	-	-	-	-	-	-

Table 3. Effects of light intensity within control and treatment groups. ▲▼($p<0.01$),▽ ($p<0.05$), - (Not significant).

FEMALES	CONTROL			EPINEPHRINE			DIAZEPAM		
	40-250	250-600	40-600	40-250	250-600	40-600	40-250	250-600	40-600
N entries	-	▼	▼	-	-	-	-	-	-
P time periphery	-	▲	▲	-	-	-	-	-	-
Dist periphery	-	-	-	-	-	-	-	-	-
Total Dist travelled	-	▽	-	-	-	-	-	-	-
Resting time	-	▲	-	-	-	-	-	-	-



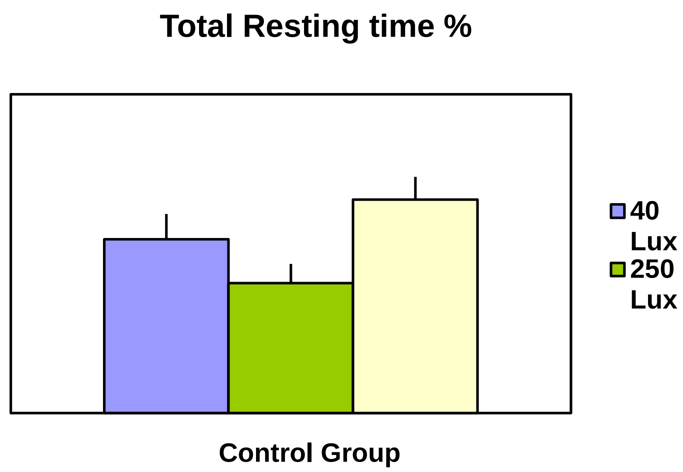


Figure 1. Total distance travelled in each light intensity (top) and total Resting Time (%) (bottom) in each light intensity in females within control group.

Within treatment groups no differences were found at any of the light intensities used, denoting the injected drugs exceeded the effects of the illumination of the arena. A slight increase in thigmotactic behavior seems to appear in males when combining anxiogenic treatment and a bright arena.

We then analyzed the same data with a different approach to try to address whether the differences in the parameters tested in treated versus untreated mice could be better appreciated at a specific light intensity, which could be extrapolated to the study of lines expected to have anxiety or exploratory related alterations (Tables 4-5).

Table 4. Comparison between treatments at a fixed light intensity. ▲▼(p<0.01), ▽△(p<0.05), - (Not significant).

MALES	40 lux	250 lux	600 lux	40 lux	250 lux	600 lux	40 lux	250 lux	600 lux
	Cont- EP	Cont- EP	Cont- EP	Cont-DZ	Cont- DZ	Cont- DZ	EP-DZ	EP-DZ	EP-DZ
N entries	▼	▼	▼	▼	▼	-	△	△	△
P time periphery	▲	▲	△	-	-	▽	▼	▽	▼
Dist periphery	▼	▼	▼	▼	▽	▽	-	-	-
Total Dist travelled	▼	▼	▼	▼	▼	-	-	-	-
Resting time	▲	▲	▲	▲	△	△	-	-	▽

Table 5. Comparison between treatments at a fixed light intensity. ▲▼(p<0.01), ▽△ (p<0.05), - (Not significant).

FEMALES	40 lux	250 lux	600 lux	40 lux	250 lux	600 lux	40 lux	250 lux	600 lux
	Cont- EP	Cont- EP	Cont- EP	Cont-DZ	Cont- DZ	Cont- DZ	EP-DZ	EP-DZ	EP-DZ
N entries	▼	▼	▼	▼	▼	▼	-	-	-
P time periphery	▲	△	-	-	△	-	▼	-	-
Dist periphery	▼	▼	▼	▼	▼	▼	-	▲	-
Total Dist travelled	▼	▼	▼	▼	▼	▼	-	△	-
Resting time	▲	▲	▲	▲	▲	▲	-	▽	-

In agreement with the results shown above, changes in female anxiety related behavior between animals stimulated with epinephrine versus control animals were dependent on illumination level. They were best seen under a dimmer light; significant differences were still present at 250 lux, but a brightly illuminated arena (600 lux) rendered statistically similar results compared to untreated animals. In males the effect was similar but less intense, supporting they may be slightly less influenced by bright arenas.

Some differences between sexes were observed when anxiolytic treated and control animals were analyzed under different brightness. Surprisingly the time spent in the periphery of the arena, considered usually as one of the main indicators of anxiety did not vary or varied in opposite direction than anticipated. Both in males and females anxiolytic treatment produced a decrease in thigmotactic behavior at 40 and 600 lux. (significantly reduced in diazepam treated males at 600 lux) as could be expected. However at 250 lux a significant increase in this parameter in females was observed; this increase could be seen in males too (although not statistically significant). These results were in agreement with a previous report on the Open Field Test in C57 mice using different doses of anxiolytic (ketamine) at 165 lux [1]. Whether said effect is specific to anxiolytic treated animals at intermediate light intensities should be confirmed.

In exploratory variables, less significant differences in distance travelled in the periphery of the arena or total resting time were seen at 250 lux in males; at 600 lux, data from diazepam treated and untreated males appeared even more uniform, while female results were unaffected.

Whether these results from exploratory and anxiety variables in males have any biological meaning (i.e. only when aversively stimulated the effect of a tranquilizing is detected by decreasing anxiety and attenuating differences in exploratory activity) or it is just an artifact of our conditions (i.e. periphery area chosen, or number of animal analyzed) should be further addressed.

The most streaking differences between male and female outcomes were observed when comparing the results of Epinephrine versus Diazepam treated animals under different illuminations. In males diazepam induced a marked reduction in the percentage of time spent in the periphery of the arena compared to epinephrine injected mice, while leaving unchanged the exploratory variables at the three lights studied. On the other hand, a very bright arena rendered statistically similar results between drug treated females in a very bright arena. However, as mentioned above, profound differences were seen when the behavior of the animals was carefully monitored. Diazepam treated female mice were calm during the test while the epinephrine injected animals were nervous and freezing behavior was present. In both cases, the “resting time” was high and the program rendered similar results. This highlights the importance of ethological analysis when performing the Open Field Test [3]. Differences in exploratory behavior between drugs were only seen at 250 Lux in females and the percentage of time spent in the periphery was only seen when analyzing data obtained at 40 lux.

Taken together our results seem to favor the use of a light intensity similar to the light the animals are exposed to, for Open Field studies. At 40 lux, all differences observed between EP or DZ treated and between treated and control animals were maximum, and almost identical between sexes. Nonetheless we cannot rule out the hypothesis that a slightly brighter arena could be better for detecting mutant lines with an expected exploratory divergence.

On the other hand there seems to be some divergences in the combined effects of light and treatment between male and female C57 mice, being the results obtained using males generally more consistent and predictable.

Finally, our results emphasize the need to perform ethological analysis when studying animals in this maze, as some values obtained in the study could become statistically similar when the animals are in fact experiencing opposite mood states.

References

[1] Akillioglu K, Melik E.B, Melik E, Boga A. (2012). Effect of ketamine on exploratory behaviour in Balb/C and C57BL/6 mice. *Pharmacology, Biochemistry and Behavior*, ***100***, 513-517.

[2] Adriaan J,Spiga F, Staub D.R.Hale M.W, Shekhar A, Lowry C.A. (2007). Differential effects of exposure to low-light or high-light open-field on anxiety-related behaviors: relationship to c-Fos expression in serotonergic and non-serotonergic neurons in the dorsal raphe nucleus. *Brain Research Bulletin*, ***72****(1)*, 32–43.

[3] Carola V, D'Olimpio F, Brunamonti E, Mangia F, Renzi P. (2002). Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. *Behavioral Brain Research*, ***134****(1-2)*, 49-57.

[4] Choleris E., Thomas A.W., Kavaliers M., Prato F.S. (2011). A detailed ethological analysis of the mouse open field test: effects of diazepam, chlordiazepoxide and an extremely low frequency pulsed magnetic field. *Neuroscience and Biobehavior Review*, ***25***, 235-260.

[5] Fraser LM, Brown RE, Hussin A, Fontana M, Whittaker A, O'Leary TP, Lederle L, Holmes A, Ramos A. (2010). Measuring anxiety and locomotion-related behaviours in mice: A new way of using old tests. *Psychopharmacology.* (Berl), ***211****(1)*, 99-112.

[6] Lizevey G, Miller J, Vogel W. (1985). Plasma norepinephrine, epinephrine and corticosterone stress response to restraint in individual male and female rats and their correlations. *Neuroscience Letters*,**62**, 51-56

How to measure episodic memory in 12-month-old infants: Implications for future research

Katarzyna Bobrowicz, Ryszard Bobrowicz

Faculty of Artes Liberales, University of Warsaw

We aim to introduce some new solutions in the studies involving the subjects who are not able to use spoken language, both animals and pre-verbal infants. Although our research focuses in general on the episodic memory introduced by Tulving in 1972, this paper is to present the case of 12-month-old infants.

Two issues should be taken into account: self-awareness and ability to form and store the individual experience [3]. They both seem to be key for the episodic memory. One may pose a simple question: is a year old infant in possession of this capacity in the light of Tulving's theory? It seems that no study has answered this question in this age group so far. These already conducted, engaged the paradigm of elicited imitation [2], while we turn to the what-where-when paradigm . Introduced by Clayton and Dickinson for the scrub jays, it emphasizes the ability to remember simultaneously three aspects of a past situation: what happened, where and when it took place [3]. If one was able to show such a recollection, it evinced episodic-like memory. In this episodic-likeness two additional Tulving's criteria are obviously missing . These are: the sense of remembering the event as the part of individual autobiography, and the self-awareness itself.

Who-where-when and self-perception

In reference to the studies with non-human primates, we propose a slight change in the Clayton-Dickinson paradigm for the infants. "Who" instead of "what" as the first aspect to check. However, the simultaneous recollection of the triad could not be satisfying, as it leave the problem of self-awareness and autoperception unsolved. Butterworth in 1992 stated that most views on the source of the concept of self in humans were based on mirror self-recognition [1]. A powder spot is placed on the infants' face, while the infant themselves is placed in front of the mirror to remove the spot using the reflection. They are able to pass the test around the fifteenth month, but the mirror may require more complex capacities than self-awareness. That is why Butterworth mentioned the Gibson's theory of direct perception, which suggests that the concept of self most likely comes from the sensory perception [1]. It has two poles here, subjective and objective, which are both described in terms of sensory stimulation and its invariant and variant properties. The first concern the invariant aspects of environment that may change in the eyes of the moving subject, who could then make distinctions between the self and non-self objects in the pole of perception. One distinction is already made by the author: between the perception and the concept of self. The latter is a cognitive representation of self based on the reflexive self-awareness. It possibly antecedes the self-perception, but we are not sure, if the lack of proof from the mirror test excludes the presence of such self-awareness.

The mirror test

Many studies refer to the mirror test. It may be traced back to 1970, when Gallup applied it to the non-human primates [1]. The mirror returns the reflected light back to its source, but its recognition is not simply the case of perception. It calls for the experience with mirrors, certain stage of cognitive development and some other conditions before self-recognition would happen. The spot is removed by the subjects circa the 15^{th} month of life, but we know that even earlier, between 8^{th} and 12^{th} month the infant may both find the objects attached to their body thanks to the mirror reflection and differentiate the contingent and non-contingent videos of themselves. We aim to link such observations with the works on *geste* by Jousse, who may once have said that a man, just like animals, thinks with his body, but is the one that is most capable of imitation among them all. We do not want to focus on the difference, but on the similarity and refer to the who-where-when paradigm.

Our study

We are to invite three groups of infants: 12-, 15-,18-month old. Three age levels are required in the light of the studies conducted previously. The study is to be conducted in an adequately adapted room with mutually connected device: a screen and a camera to record the behavior of three people in the experimental space: the infant, their caretaker and the experimenter. In the third phase a chair would be included in the space as well, so the caretaker could sit with their infant and watch the material on the screen. The study is to include four phases.

1. In the first phase there would be the caretaker, the infant and the experimenter. Real time video from the camera would be displayed on the screen. If the infant was not interested with it in the first five minutes, the experimenter would afterwards point his finger on the screen and say "Look!". Whole phase would last 10 minutes.

2. The second phase is to start immediately after the previous one. An absorbing toy would be then introduced, moving and emitting a sound. For the next 5 minutes the infant could play with it and simultaneously watch the current situation on the screen.

3. The test phase is to take place 24 hours later. The infant would enter the same space; but, the camera would be placed above the screen to record what and how long the infant looks at, sitting on the caretaker's knees. Two series of three sets of videos would be displayed. Series 1: In the first set, on the left side of the screen a genuine video would be placed, on the right though – a modified video with a Photoshop-substituted toy to verify the aspect of what. In the second set, the space (where). In the third set, the experimenter (who) are to be substituted in a similar way. Series 2: the sides would be reversed.

4. An additional trial should be introduced for each child afterwards: the mirror test itself to check, whether they were able to pass the mirror test.

The final remarks

Although we expect that in all age groups the time of looking at the modified videos should be significantly longer, we have not conducted our study yet. We expect that 18-month-olds would generally pass the mirror test and focus their sight on the modified videos significantly longer than on the non-modified. 15-month-olds, as some of them are claimed to pass the mirror test, should also generally evince this pattern, though less often. 12-month-olds on the other hand should not evince it according to the previous studies, but we expect them to do so, regardless of their result in the mirror test. This would engage the habituation-dishabituation paradigm, which suggests that the child should look significantly longer for unfamiliar stimuli in comparison with the familiar.

The study is to be carried out mostly in 2015, however we would like to discuss our methods to apply them in the future to even younger infants and, possibly, non-human primates. We are aware of the fact that studies on episodic memory were conducted with the participation of 3- and 4-year-old children, but we would like to develop methods without the use of language and need of verbalization.

Ethical statement

All children and parent to be involved in our study will be taken care of. The study will be of a voluntary character and shall be ceased once the child and/or their parent does not want to participate any longer. A feedback will also be provided for the parents on the results and general performance of their child.

References

[1] Butterworth, G. (1992). Origins of self-perception in infancy. *Psychological Inquiry, **3***, 103-111.
[2] Fivush R. (2011). The development of autobiographical memory. *Annual Review of Psychology, **62***, 559-582.
[3] Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, **53***, 1-25.

How to measure episodic-like memory in rats and cats: Implications for future research

Ryszard Bobrowicz, Katarzyna Bobrowicz

Faculty of Artes Liberales, University of Warsaw

We aim to introduce some new solutions in the studies involving the subjects who are not able to use spoken language, both animals and pre-verbal infants. Although our research focuses in general on the episodic memory introduced by Tulving in 1972, this paper is to present the case of Wistar rats and domestic cats.

What-where-when

Tulving introduced the term of episodic memory as the ability to recall facts and events awarely. Episodic memory was to record what, where and when happened, but semantic memory was to do this as well. However, the episodic element meant that the individual was able to re-experience the events from their autobiography. The human subjects evinced such an ability in verbal interactions outside the experimental space, which was tough in case of the non-verbal subjects. That is why in 2003 Clayton and Dickinson proposed the term of episodic-like memory for integrated recall of past events in case of what-where-when [2]. In addition, it should be elastic so that the animal may use their memories under the new circumstances. The concept was then introduced into the studies by Eacott and Norman [3], Babb and Crystal [1] and Kart-Teke et al [6]. In 2012 we decided to replicate the last study as it claimed to prove episodic-like memory in rats.

The procedures concerning the “what” aspect include two phases: one training with the exploration of two copies of object A and one test, where the B object is introduced instead of one copy of A. B should be explored longer than A. The “where” aspect includes training and test again: first a copy of A is introduced in certain location and then second copy of A is placed in new location, expected to be explored significantly longer. Finally, the “when” question is answered thanks to two training phases and a test one. In the first training a copy of A and a copy of B are presented, in the second: just A. In the test A and B are presented again and the subject should explore B longer.

Kart-Teke et al.

The study by Kart-Teke et al. integrated these procedures in a way [6]. It included three phases: two training and a test one. Each lasted for 5 minutes and in each four copies were placed in four different locations. All objects had the same material with different shape, size, colour and surface. The object were heavy enough so that the rats could not move them. There were two types of objects chosen randomly for each rat. In the first training the subject was placed in the open field and presented with four copies of object A in four different locations. After the 50-min pause four copies of B were presented in four other locations. After second 50-min pause the test phase started. One copy of A was placed in its previous location and another copy of A was dislocated. Likewise, one B was placed as previously and another was dislocated. The objects were explored if the rat touched it with their nose or legs. The authors expected them to explore the earlier, non-dislocated object significantly longer than the later non-dislocated object and to spent more time exploring the dislocated than the non-dislocated objects. Then they would evince the recall of “where” and “when”. These correlations proved significant along with another: the rats also preferred the later non-dislocated objects to the earlier dislocated. Altogether, according to the authors, they indeed seemed to evince episodic-like memory.

Our pilot study

These conclusions inspired us to conduct a pilot study to replicate the procedure with a few changes. First, we introduced an eight-arm maze instead of the open field. Secondly, we introduced three types of wooden objects of different tone, shape and surface which could have been easily moved. Each of three phases included 10 trials of 15 min. They were conducted twice a day: in the morning and in the afternoon. A 6-day pause between the second training and the test was introduced. We chose randomly a set of three objects: A, B and C for each of the ten Wistar rats: five male and five female. Two copies of A were presented in the first training in two personalized arms, and two copies of B in the second training in two other arms. In the test a copy of A and a copy of B were introduced in the same arms as previously, while one copy of each was dislocated. Also two copies of C were introduced in two out of four remaining arms. In the aspect of "when" we expected that the pair of A objects would be explored significantly longer than the pair of B objects. In the aspect of "what" we assumed that the rats would spent more time exploring the C objects, and in the aspect of "where": more time exploring the dislocated than non-dislocated objects.

Main study

The results failed to prove our expectations. There was no significant relationship neither between the exploration and the object type nor its dislocation. We considered the objects too alike and decided to introduce more distinctive objects: a wave-shaped porcelain white cylinder, a mushroom-shaped metal grey object and a standard wooden light brown cuboid. The trials were shorter and lasted for 10 min, and there was no pause between the training phases and the test phase. We also involved 10 male subjects instead of 4 male and 4 female that participated in the pilot study.

During the testing the maze was positioned 20 cm above the ground to prevent the animals from the escape and, simultaneously, fit the view of the video camera placed above the maze. White light of 20 lx iluminance was introduced due to the shape variation of the objects presented. The videos were watched one by one with no use of any specialized programme and the timing was measured with a stopwatch.

As expected, the rats spent significantly more time exploring the new than the old , both A and B, objects (BCI: 1.86158-7.25381 and 2.46398-7.49567, correspondingly). They also explored C object significantly more often than the familiar objects regardless of their dislocation (BCI: 0.124-0.940). It suggested that the rats did recognize both the "what" and the "when" aspect of the appearing objects. However, we failed to check their awareness of the "where" aspect, as the difference between the stationary A and dislocated A was not significant. It was similar for the stationary B and dislocated B, however in both cases the insignificant tendency was observed. Its weakness was probably caused by the number of the test trials and – consequently – the number of variables. As the stimuli seem to be adequate, more trials, both training and test, should be conducted in the future to prove the hypothesized differences significant.

Systematic observation with cats

We decided to apply our procedure and background to cats with a key change: form the "what" aspect to "who". We have planned pilot observation, where two male cats and a female were introduced into the modified open field covered with a thick blanket in a room with dim light. We used a box with an entrance cut out in the shorter side and connected with a start box with a sliding gate. Two additional exits were made in the farthest ends of the perpendicular walls, both with similar sliding gates. Each cat received twenty four training phases and four test phases, which covered all possible sets of who-where-when. Each phase consisted of two parts – exploration and choice. In first part of the training the animal was allowed by one of the experimenters into the start box for a second. Then the gate went up allowing into the field, where both gates remained open and were explored for thirty seconds. Right after, a one- or fifteen-minute delay was introduced and then the cat was allowed to the field again with one open gate only, either left or right according to the procedure. Once the cat used it to leave the field, the phase was over. We expected that in the test phase they would go first to the closed gate when given the choice and then to the open one. We could observe their behavior thanks to the motion under the blanket. Each set of who-where-when was individualized and random.

The results seem to prove the episodic-like memory in the subjects. The older male responded properly in all four test phases, the younger male in 3 out of 4, and the female failed the test due to her preference for one gate.

Implications for future research

We would like to discuss our methods to apply them to other species if possible. We have also some ideas for their enhancement, which we are willing to discuss during the presentation.

Ethical statement

We conducted our rat studies thanks to the permission of First Warsaw Local Ethics Committee for Animal Experimentation. The permission of the main study received the following number: 369/2012. Our cat observation involved the cats we take care of and was of a play character for them. They were not forced to take part in our project. All studies focused on animal behaviour and neither involved any invasive interaction nor caused any discomfort to the animals.

References

[1] Babb S.J., Crystal J.D. (2005). Discrimination of what, when, and where: Implications for episodic-like memory in rats. *Learning and Motivation*, ***36***, 177 – 189.

[2] Clayton, N. S., Bussey, T. J., Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience*, ***4***, 685-691.

[3] Eacott M.J., Norman G. (2004). Integrated Memory for Object, Place, and Context in Rats: A Possible Model of Episodic-Like Memory? *The Journal of Neuroscience*, ***24(8)***, 1948 – 1953.

[4] Gaffan D. (1994). Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. *Journal of Cognitive Neuroscience*, ***6***, 305–320.

[5] Hampton R.R., Schwartz B.L. (2004). Episodic memory in nonhumans: what, and where, is when? *Current Opinion in Neurobiology*, ***14***, 192 – 197.

[6] Kart–Teke E., De Souza Silva M.A., Huston J.P., Dere E. (2006). Wistar rats show episodic-like memory for unique experiences. *Neurobiology of Learning and Memory*, ***85***, 173 – 182.

[7] Squire L.R., Zola S.M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences,* ***93***, 13515-13522.

[8] Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology,* ***53***, 1-25.

Local Field Potential:

A link between behavior and physiology

Marta U. Wołoszynowska-Fraser, Gernot Riedel

Institute of Medical Sciences, University of Aberdeen, Aberdeen, Scotland, United Kingdom.
r01muw11@abdn.ac.uk

Introduction

Cognitive events bind ensembles of neurons together to bring about a uniform behavioral response. The elements of these ensembles are typically neurons with their own individual properties encoding/retrieving specific task parameters so that an overall appropriate response pattern can be observed. Yet, the same neurons may exhibit differential neuronal activity in a spatial versus a non-spatial task, or a short-term relative to a long-term memory task. Characteristic firing patterns have been recorded from extracellular electrodes and, at least for hippocampus, theta oscillations can be demodulated and reveal the spatial location through feature-tuned field potentials [1]. More globally, theta oscillations are associated with the complex behaviors, including processing of memory and decision making, and working memory and navigation in rodents and humans [2] and their propagation from hippocampus to prefrontal cortex can entrain prefrontal gamma oscillations to activate attentional processes, cognition and working memory [3,4]. These rhythms are typically recorded in animals by multi-electrode arrays recording several dozens of cells and are locally restricted to the few neurons in the proximity of the electrode.

This is in stark contrast to human studies, in which brain activity is readily accessed through scalp electrodes placed in the outside of the scalp in regular rows on predetermined spaces in a high density of sometimes 256 probes constituting the EEG. At the same time, the human test subject is required to perform specific tasks so much so that the EEG can be fragmented to carry trial-specific information relative to stimulus onset, inter-trial interval, or recall phase. Such an approach is widely lacking in animals, particularly mice given the size and fragility of the skull, and the finesse required to align behavioral stimuli and responses with EEG measures with high fidelity. Here we describe the development and implementation of such an approach using a light-weight wireless EEG device connected to the head of the mouse in a working memory paradigm.

Methods

All experiments were performed under the United Kingdom Animals (Scientific Procedures) Act 1986 and EU directive 63/2010EC and approved by the local Ethical Review Committee. Mice (C57BL/6, locally bred, both genders, 8-10 weeks old, no water or food restrictions) were stereotaxically implanted with bilateral gold electrodes placed into prefrontal cortex and hippocampus. For recording of continuous local field potentials, NeuroLoggers (NewBehaviour, Zurich, Switzerland) were attached to the head-stage before mice entered the behavioral apparatus. To measure working memory we recorded spontaneous alternation in the Y-maze. Using EthoVision 3.1 (Noldus Information Technology, Netherlands) we defined zones in the Y-maze: distal zones (A1, B1 and C1) and proximal zones (A2, B2 and C2). Entry into these zones generated TTL pulses, which were transmitted via a buffer to the infrared (IR) pattern generator (Fig.1). Each zone entry was identified by a different IR pattern and recognized by the IR sensor on the NeuroLogger. IR signals were aligned to LFPs, downloaded and processed with MatLab® together with the neuronal activity and analyzed using BrainVision Analyzer 2.0 (Brain Products GmbH, Germany). Recording were segmented according to the IR stamps with the view to determine neuronal differences when animals pass through or explore different zones of the maze (spatial sorting). Extracted episodes of brain activity were then Fourier transformed and compared for absolute and normalized power. Statistical comparison used parametric tests with alpha set to 5%.

Results

The ability to pinpoint the brain activity in the regions of interest to a particular place in the maze helps to understand alterations in oscillatory activity in response to different cognitive demands. In theory, one could argue that the distal zones are regions of pure exploration with very little valence; at the same time, all distal zones are proposed to be of equal relevance. In contrast, the proximal zones carry higher cognitive load as a decision has to be rendered whether to turn left or right (or re-enter the same arm). Subjects displayed lower prefrontal power (normalized) in the decision (proximal) zone compared with the distal zone in the all frequency bands analyzed including theta, alpha, beta and gamma (low, high) (F's>5; p's<0.02). The hippocampal normalized power in the proximal zones was lower only in 20-50Hz range (p=0.0009, Df=1, F=11.28) compared with the distal zones.

Conclusion

We here present a novel means for the alignment of behavioural observations and wireless recordings of oscillatory activity in the brain. This was established in a Y-maze paradigms with the particular view to recording of working memory, but we have also utilised the method in other behavioral settings (open field, Barnes maze, etc...). In this experiment, animals readily modulated their neuronal activity depending on cognitive load. Overall, they lowered the power of LFPs when passing through the decision zones/heightened their activity when exploring the distal zones. This ability to pinpoint the activity of neuronal ensembles in brain regions of interest to specific cognitive functions thus allows a translational approach matching the measurment of EEG changes during cognitive testiing in human controls and patients.

References

[1] Agarwal, G., Stevenson, I.H., Berenyi, A., Mizuseki, K., Buzsaki, G., Sommer, F.T. (2014) Spatially distributed local fields in the hippocampus encode rat position, *Science*, ***344***, 626-630.

[2] Benchenane, K., Tiesinga, P.H., Battaglia, F.P. (2011) Oscillations in the prefrontal cortex: a gateway to memory and attention. *Current Opinion in Neurobiology*, ***21***, 475-485.

[3] Howard, M.W., Rizzuto, D.S., Caplan, J.B., Madsen, J.R., Lisman, J., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., Kahana, M.J. (2003) Gamma oscillations correlate with working memory load in humans. *Cerebral cortex*, ***13***, 1369-1374.

[4] Povysheva, N.V., Zaitsev, A.V., Rotaru, D.C., Gonzalez-Burgos, G., Lewis, D.A., Krimer, L.S. (2008) Parvalbumin-positive basket interneurons in monkey and rat prefrontal cortex. *Journal of Neurophysiology*, ***100***, 2348-2360.

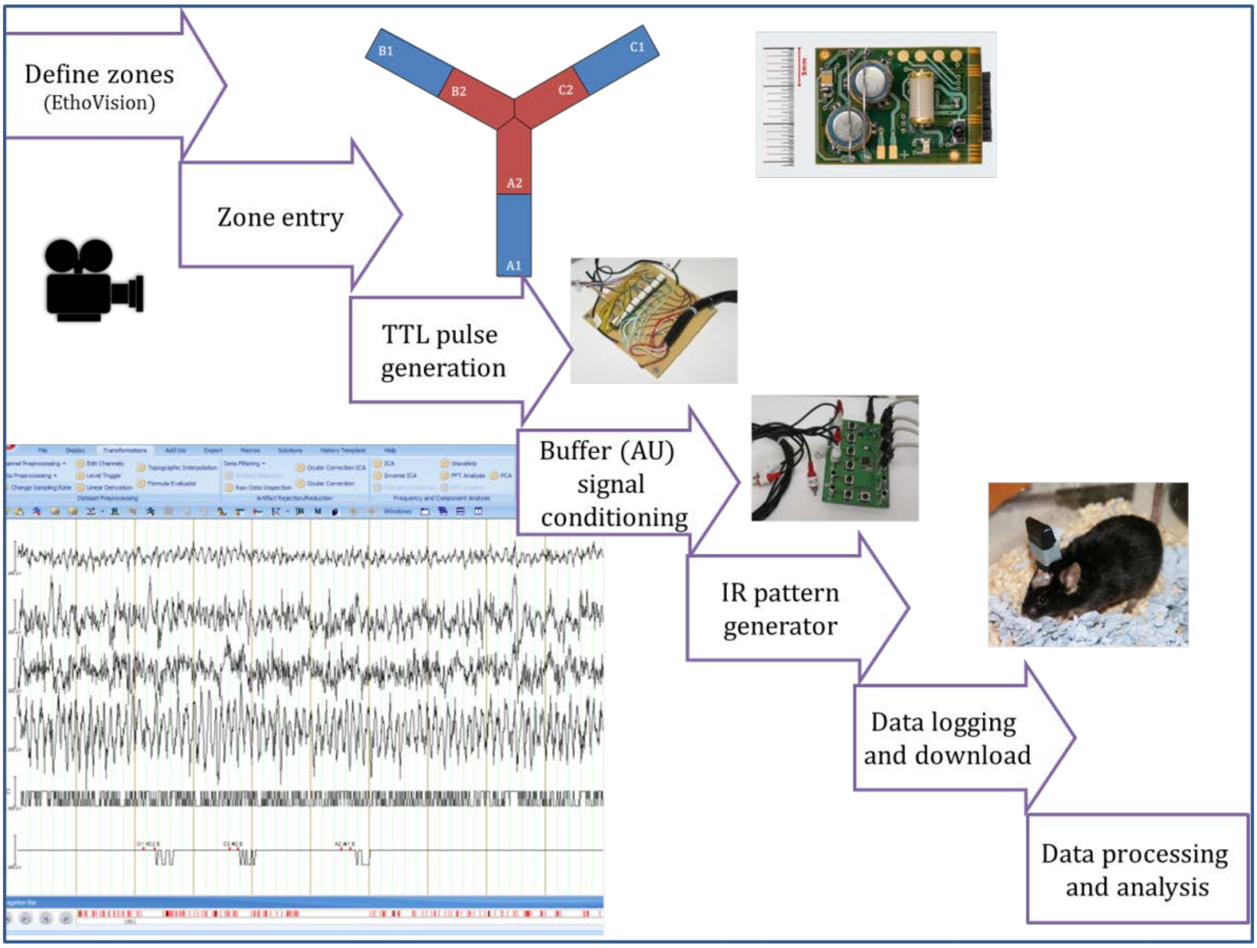


Figure 1. Behavioral and EEG setup and analysis. Outline of stepwise alignment between spatial behavior and physiological observation using a zonal arrangement in the Y-maze (A1-C2) combined with wireless NeuroLogger recording in freely moving mice. For details see text.

Sound Analysis of Dairy Cows

G.H. Meen[1], M.A. Schellekens[1], M.H.M. Slegers[1], N.L.G. Leenders[1], E. van Erp-van der Kooij[1], and L.P.J.J. Noldus[2]

[1]Department of Applied Biology, HAS Den Bosch University of Applied Sciences, ‘s-Hertogenbosch, the Netherlands
[1]L.vErp@has.nl
[2]Noldus Information Technology B.V., Wageningen, the Netherlands

The ambition of modern farming is to have a productive, healthy and happy livestock. To achieve this, there is a growing demand for information about the individual animal as well as the group. However, because of the growing number of livestock per farm there is less time for the farmer to observe each individual animal [1, 6].

Precision Livestock Farming (PLF) supports the farmer in giving each individual animal sufficient attention. PLF is technology which is used to continuously measure parameters such as activity, feed intake or oestrus activity. Examples of PLF using sound sensors are found in pig farms. In 2001 Vandermeulen et al. examined the correlation between image and sound recordings from pigs. He concluded that the use of video and audio data has potential as a tool to improve livestock welfare. The use of sensors can facilitate monitoring large livestock groups [5].

The use of sound analysis in cows was studied by Temple Grandin [3]. She showed that cows vocalised significantly less when they were not pushed with electrical sticks to the place where they got slaughtered. This may indicate that the mooing is generated by stress or pain caused by the electrical stick. It is interesting to investigate whether the behaviour of a cow can be predicted by their uttered sounds. Noise analysis could be used as an early warning tool, for example to detect anxiety or oestrus.

In this study, we recorded sound and behaviour from dairy cows by audio and video recordings. Four cameras and microphones were placed in a modern dairy farm in Herwijnen, The Netherlands. Three cameras and microphones recorded dairy cows, one camera and microphone recorded heifers between four and ten months of age. First a 5-days period of 10 recording hours per day was used to test the technical installation and to adjust microphones and cameras. Dampers were installed for noise reduction. Next, recordings were made for 17 days, 10 hours per day. Two days of recordings were discarded due to technical problems, and the remaining 15 days of 10 hours of synchronised audio and video recordings were analysed.

After the background noise was filtered out, the calls of cows were traced and if possible linked with the simultaneously expressed behaviour. The behaviour was determined with the use of an ethogram. Video analysis was conducted using The Observer XT 11.5 (Noldus Information Technology B.V.). The calls were classified into different groups based on frequency, amplitude and wavelength using UltraVox 3.0 (Noldus Information Technology B.V.). Statistical analysis will demonstrate whether there is a significant resemblance between the classified groups and behaviours.

This study is the first step towards using sound analysis as a tool for dairy cattle management. If behaviour can be predicted with sound sensors, dairy farmers can be alerted quickly when stress calls are detected. Welfare, health and growth of young cattle is strongly influenced by stress [2, 4]. With the use of sound detectors, management of these problems will become more easy.

References

[1] Banhazi, T.M., Lehr, H., Black, J.L., Crabtree, H., Schofield, P., Tscharke, M., Berckmans, D. (2012). Precision Livestock Farming: An international review of scientific and commercial aspects. *International Journal of Agricultural & Biological Engineering,* ***5***, 1-9.

[2] Collier, R. J., Beede, D.K., Thatcher, W.W., Israel, L.A., Wilcox, C.J. (1982). Influences of Environment and Its Modification on Dairy Animal Health and Production. *Journal of Dairy Science,* ***66***, 2213-2227.

[3] Grandin, T. (1998). The feasibility of using vocalization scoring as an indicator of poor welfare during cattle slaughter. *Applied Animal Behaviour Science,* ***56***, 121-128.

[4] Heinrichs, A.J. (1993). Raising Dairy Replacements to Meet the Needs of the 21st Century. *Journal of Dairy Science,* ***76***, 3179-3187

[5] Lokhorst, K., Fels, van der, B., Riel van, J., Hogewerf, P., Holster, H., Lourens, S. (2011). Verkenning High Tech Diermanagement in de varkens en pluimveehouderij. *Wageningen UR livestock Research.*

[6] Lokhorst, K., Ipema, B., Bleumer, E., (2010). Haalbaarheid precisielandbouw voor de melkveehouderij. *Wageningen UR livestock Research.*

[7] Vandermeulen, J., Kashiha, M., Ott, S., Bahr, C., Moons, C., Tuyttens, F., Niewold, T., Berckmans, D. (2001). Combination of image and sound analysis for behaviour monitoring in pigs. *Precision livestock farmin*, ***2013***, 263-268.

New media supported, child-oriented Pilates as intervention to stabilize posture and to correct postural defects on pupils aged 10 – 12

D. Deutschmann[1], M. Parfant[2], H. Holzer[3], M. Ebner[4], G. Ivanic[5], M. Svehlik[6], A. Weiglein[7]

[1]Medical University Graz, Austria.
[1] dietlind.deutschmann@stud.medunigraz.at
[2]Graz University of Technology, Austria.
[2] m.parfant@student.tugraz.at
[3]Institute of Sports Science, KF University Graz, Austria.
[3] hans-peter.holzer@uni-graz.at
[4]Institute for Information Systems and Computer Media, Graz University of Technology, Austria.
[4] martin.ebner@tugraz.at
[5]Department of Orthopedics, Privatklinik Graz Ragnitz, Austria.
[5] gerd.ivanic@ok-institut.at
[6]Department of Pediatric Surgery, Medical University Graz.
[6] martin.svehlik@medunigraz.at
[7]Institute of Anatomy, Medical University Graz, Austria.
[7] andreas.weiglein@medunigraz.at

Introduction

Nowadays back pain due to postural defects is one of the most frequent chronic symptoms in our population. Particularly the number of school children with postural defects is increasing [1]. The earlier a weakness in posture, influenced by muscular imbalance for example, is recognized the higher the chances to prevent further defects [2]. The aim of this study is to show the effect of a regular short Pilates-program to stabilize posture in school children. New digital media is supposed to support subjects' motivation and help evaluating their activity.

Methods

For this study 128 pupils (46 males and 82 females) at an average age of 10,77 (SD 0,30) years were recruited. The study has been conducted in two secondary schools in Graz and nearby Graz. Posture was tested by the spine analysis program Zebris CMS-System [3] and muscle function was measured using the tests of Janda [4].In addition, a questionnaire on physical activity based on the MOMO-questionnaire [5] was implemented. Subjects were divided into intervention (n=77) and control groups (n=51) (see Tabel 1).

Table 1. Demographic data of intervention and control group at the beginning of the study.

	Intervention group (n=77)	Control group (n=51)
age [years]	10.77 ± 0.32	10.78 ± 0.27
sex (m/f)	37 (48%) / 40 (52%)	9 (18%) / 42 (82%)
residence (urban/rural)	27 (35%) / 49 (63%)	12 (24%) / 39 (76%)
height [cm]	147.00 ± 7.67	145.69 ± 6.98
weight [kg]	37.53 ± 7.08	37.90 ± 7.05
percentile [%]	45.86 ± 30.48	51.05 ± 27.81
sports activities (frequency/week)	4.72 ± 1.41	4.24 ± 1.51

The intervention took place in sports lessons at school over a total period of one year and two months. Children in intervention group should also perform the sports program – ten exercises with ten repetitions in ten minutes with the help of the little magician "*Wirakulix*" – at home every day but at least three times a week. Subjects in control group did not get any intervention, they just participated at the regular sports lessons in school.

For monitoring and supporting this task, an iPhone-application, which is called *MotionTracker,* was especially implemented for this study (Apple AppStore [6]). All participants had to register their movements by the help of a diary, an online-diary or the iPhone- app. 17 out of 77 kids (22,1 %) had the appropriate advice to use the app. iPhones are shipped with an accelerometer and a gyroscope. Both sensors supply data for the three-dimensional space with a sampling rate of 50 hertz. Figure 1 shows the rotation rate over the time, which is delivered from the gyroscope sensor in the iPhone, of the exercise “Spine-Twist”. The red curve corresponds to the rotation rate in the z-axis, which has a significant amplitude in this movement. On the other hand, the rotation rates of the green x-axis and the blue y-axis are relatively small.

The provided data was used for classification. A typical classifier consists of five components - collection, segmentation, feature extraction, classification and post processing [7]. The segmentation was done with peak detection on the gyroscope data. Each exercise had a significant shaping which was utilized. Using these extracted gained data sets, a data matrix which consisted out of six columns was built. These columns were the three-dimensional rotation and acceleration data and the count of rows depended on the time interval. From this resulting matrix different features were extracted for classification with support vector machines. Li, Kulkarni and Prabhakaran suggest to use singular values for classifying motion patterns [8]. These singular values as well as classical features like the mean (expected value) and the standard deviation from each of the different dimensions, cross-correlations between them and between the whole three-dimensional acceleration and rotation data, were used.

The different categories were interpreted as exercises in this context. Support vector machines divide two categories with a so called hyperplane that maximizes the distance between them. To calculate this optimal hyperplane training data is needed, the calculation and optimization is called training of the classifier. This training data was produced using different kids as test subjects. They had to perform the different exercises and also the different typical errors. For this purpose thirty kids performed a total of 451 repetitions of the different exercises, which means that each exercise had an average of about 75 positive training samples. Training samples of different exercises as well as simulated samples served as negative training samples. The relation of positive to negative samples and subcategories (e.g. typical error, well done) to each other was about 1 to 1.

For each exercise a fixed reference point was defined, which is fundamental for its movement and to classify the gained data from the sensors of the iPhone. Before the kids started with their workout, they had to attach the device to this reference point (see Fig. 2). When they had finished the sports program, which was signalized by a simple touch of a button, they got the evaluation of the recorded movements. *MotionTracker* counts repetitions, provides feedback about the accuracy and suggestions for improvement of the accomplished exercises.

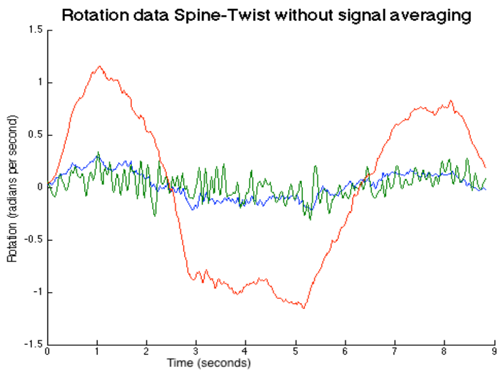


Figure 1. Rotation data, delivered from the gyroscope sensor of the iPhone, of the exercise “Spine-Twist”

Figure 2. A kid with an attached device during workout

The whole workout is stored on a server and evaluations of the workout history can be found by the kids and also the therapist via a simple web application. Furthermore, the therapist gets additional information about the rate of participation and precision of the movements. Finally after more than one year of intervention all tests from the beginning of the study were repeated.

Results & Discussion

Basically, the app enables a possibility to perform exercises at home in a controlled situation. Although this approach provides good results, it is somehow limited, as it can only monitor one fundamental part of the movement. Because of that, not all of the movements can be monitored, i.e. legs cannot be monitored when the reference point is on one of the arms. The implemented *MotionTracker*-system was also evaluated with four children, all with different prerequisites concerning the age and the physical level. They were told to accomplish ten repetitions of all six exercises of the workout, which could be registered with the phone. Practically this was not always possible because the children performed test repetitions, counted wrong or were not able to finish it because the fitness level was too low. Altogether there were found 633 possible repetitions of the different exercises, 223 were real repetitions and 410 randomly found segments. The implemented system could recognize 189 of the 223 real repetitions, which resulted in a hit rate of about 84.75 percent. 73 of the 410 randomly found segments were falsely assigned to real repetitions; this corresponded to a fallout rate of about 17.8 percent. Totally the correct classification rate is about 83.1 percent.

The result showed that nearly all real repetitions could be found. Relations between the complexity of the exercises, the amount of movement in it and the hit rate were noticed. More complex exercises achieved the lowest hit rates. For all others a practically sufficient hit rate of more than 95 percent could be achieved. The next essential step is to improve the rate of participation through gamification. This means motivational elements must be added to increase the fun factor for the kids. The analysis of the intervention is still going on but first results show that in intervention (IG) as well as in control group (CG) muscle stretching decreased (IG: $p=0{,}002$; CG: $p=0{,}001$) and muscle strength increased (IG: $p=0{,}001$; CG: $p=0{,}021$) significantly. In observation of the intervention group separated in subjects who did fulfil the exercise recommendations (26,8 %) and in those who did not (73,2 %), it is recognizable, that subjects who practised less than three times a week had a significant decreasing in muscle stretching ($p=0{,}042$) and no significant change in muscle strength. At the other hand subjects who fulfil recommendations did not significantly change in muscle stretching but increased their muscle strength significantly ($p=0{,}014$). It is recognizable that posture of the intervention group was better at the end of the study. However, not all of the results concerning the effect of the sports program on the parameters of posture have been analysed yet.

References

[1] Latalski, M., Bylina, J., Fatyga, M., Repko, M., Filipovic, M., Jarosz, M.J., u. a. (2013). Risk factors of postural defects in children at school age. *Annals of Agricultural and Environmental Medicine*, ***20****(3)*, 583–587.

[2] Newcomer, K., Sinaki, M., Wollan, P.C. (1979). Physical activity and four-year development of back strength in children. *American Journal of Physical Medicine & Rehabilitation*, ***76***, 52–58.

[3] Küster, M. (2003). Dreidimensionale Ultraschalltopometrie der Wirbelsäule und Maximalkraftmessung der Rumpfmuskulatur bei Jugendlichen. *Deutsche Zeitschrift für Sportmedizin*, ***54****(12)*, 352–354.

[4] Janda, V. (2000). *Manuelle Muskelfunktionsdiagnostik*. 4. Auflage. München, Jena: Urban & Fischer.

[5] Bös, K., Worth, A., Opper, E., Oberger, J., Romahn, N., Wagner, M., u. a. (2009). Motorik Modul. Eine Studie zur motorischen Leistungsfähigkeit und körperlich-sportlichen Aktivität von Kindern und Jugendlichen in Deutschland. Baden-Baden.

[6] Apple Inc. (2013). MotionTracker. <https://itunes.apple.com/at/app/motiontracker/id732787889?mt=8>. Last accessed on 18 March 2014.

[7] Duda, R.O., Hart, P.E. & Stork, D.G. (2000). *Pattern Classification (2nd Ed.)*. Wiley-Interscience.

[8] Li, C., Kulkarni, P.R. & Prabhakaran, B. (2007). Segmentation and recognition of motion capture data stream by classification. *Multimedia Tools and Applications.*

Ethical Statement

The study was approved by the ethical committee of the Medical University Graz, Austria (15.02.2013, EK-Number: 25-130 ex 12/13)

New approach to analyse navigational search strategies used by mice during a water maze task

P-H. Moreau[1,2,3], V. Melis[1,2,3,] B. Crouch[1], C. Harrington[1,2], G. Riedel[1]

[1]School of Medicine and Dentistry,University of Aberdeen, Aberdeen, Scotland, United Kingdom
[2]TauRx Therapeutics Ltd., Singapore.
[3]pierre-henri.moreau@abdn.ac.uk
v.melis@abdn.ac.uk

The open-field water maze task is one of the most widely used behavioural tasks to investigate cognition and to assess spatial learning and memory in rodents [1]. The principle of the task relies on the motivation to locate a fixed hidden platform as the only mean of escape from the water. An improvement in performance recorded over trials and days, for example of latency or path length to reach the platform location, reflects learning and is typically expressed as a reduction in the length of the swim path. During recent years, in addition to the analysis of common proxies (path length, swim speed, latency to platform), many researchers have shown a growing interest in the qualitative aspect of learning. Indeed, it is now established that in the water maze the animal develops navigational search pattern based on the use of multiple different strategies [2]. As a consequence, analysis of these strategies would be helpful, for example, in discriminating between groups of animals that show otherwise no differences according to the standard parameters. In light of this consideration, investigation of the type of strategy used to solve the task provides a more refined and comprehensive analysis of spatial learning in the water maze.
Search paths, such as those described by Garthe and colleagues [3], are now categorised into 3 main classes: "*Spatial, Procedural and Random*". The spatial search strategies are commonly classified as "*Directed, Focal* and *Direct"* and the procedural ones as "*Scanning and Chaining*" (see Figure 1). In the literature, most analyses are performed visually through a process of manual scoring. However this method can be time consuming and may lead to biased interpretation of results linked to the subjectivity of the experimenter.

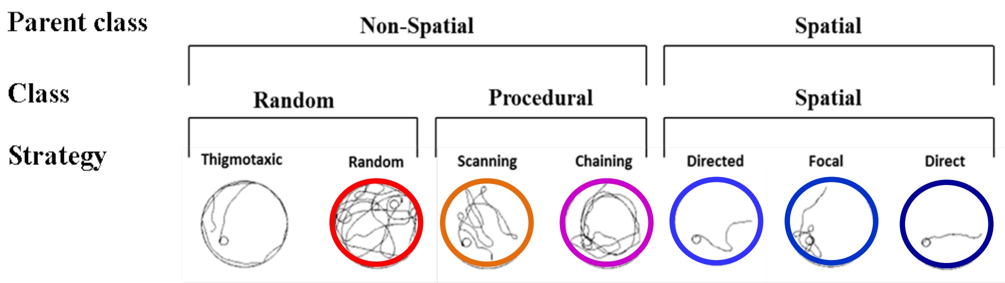


Figure 1. General classification of the search strategies commonly used in the water-maze. Parent classes are defined as *Spatial* and *Non-spatial*. They further dissociate into 3 classes, within the non-spatial are regrouped *Random*: path with no directional preference, *Procedural* defined as an allocentric search and *Spatial* related to a search based on the use of visual landmarks. The individual members in these classes include: *Thigmotaxis*: "Wall hugging", *Random*: total pool surface covered with no directional preference, *Scanning*: scanning of environment confined to the centre of the pool, *Chaining*: circular swimming at a fixed distance of the pool, *Directed*: swim with a directional preference, *Focal*: search close to the platform, *Direct*: direct navigation to the platform. Learning typically progresses through these categories from random searches to direct swims.

The aim of this study was to provide a more detailed analysis of the water maze by means of a new automated strategy analyser developed in our group and to assess its strength in discriminating search strategies relative to manual scoring. The MATLAB script is based on the highly flexible fuzzy logic sorting system, which classifies swim tracks automatically into episodes corresponding to search strategies (see Crouch et al. in this conference for details of methods).

In this experiment, two different transgenic mouse lines, named Line 1 and Line 66 aged 6 and 3 months respectively were trained in the open-field water maze task and search strategies were analysed both visually and automatically. Mice were housed in groups of up to 10 and allowed food and water *ad libitum* and were kept under standard conditions (temperature 20−21°C, 60−65% relative humidity) on a 12 hours light/dark cycle (light on at 7:00 a.m.). Tests took place during the light phase of the cycle. All experiments were conducted with local ethical permission and in strict accordance with UK Home Office regulations outlined in the Animals (Scientific Procedures) Act 1986.

Learning in the water maze was recorded as a progressive lowering of the distance moved to locate the platform, no difference was found for Line 66 and wild-type mice and both groups improved their performance over days (see Figure 2A). Similarly, the visual manual scoring, as well as the automatic analysis, did not show differences in the pattern of strategies used by the mice during the task. From day 1 to day 4 of training, both analysis systems highlighted a progressive diminution in random searches while increasing the propensity of spatial strategies (see Figures 2B, C). By contrast, Line 1 animals present with impairment in learning the water maze task as indicated by a longer distance moved to reach the platform compared the control mice (see Figure 3A). This deficit was confirmed with both strategy analyses (see Figures 3B, C) revealing a persistent level of random searches along with a lower percentage of spatial strategies on day 4.
Regarding the comparison between the visual and MATLAB script analyses, the automatic scoring matches 92% of the visual analysis in terms of spatial and non-spatial parent classes, but fell to 60% in terms of sub-divisions strategies. This can be seen for example for the search strategies of control animals recorded during the first day of both experiments, where the percentage of scanning reaches up to 70% when visually analysed and less than 10% in F-logic. On the contrary, the level of random and chaining strategies detected automatically is higher than the visual manual scoring. This difference could be explained by a different classification method (see Crouch et al., this conference). Indeed, the experimenter can only base the analysis on the entire track which leads to only one possible choice of strategy. By contrast, F-logic segments the single track into different episodes assigning the appropriate strategy based on the probabilities of occurrence and selecting the strategy with highest probability value.

In light of these results, the use of an F-logic based analysis reduces the subjective characterisation of search strategies attributed with a visual manual scoring and offers more flexibility in the qualitative assessment of phenotyping spatial navigation.

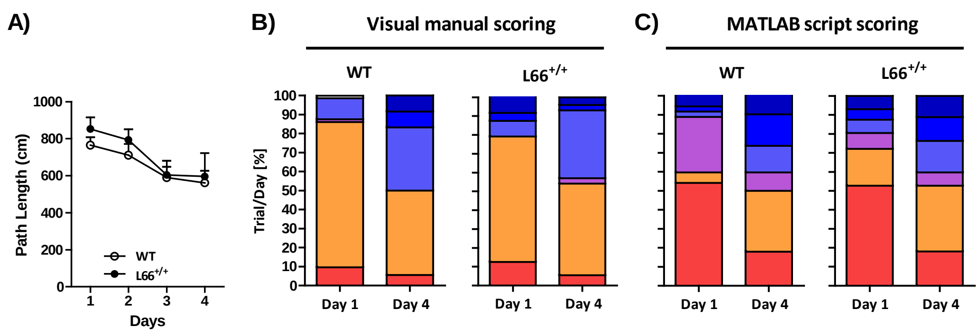


Figure 2. Comparison between visual and automatic analysis regarding the strategy used by transgenic mice (L66$^{+/+}$) and wild type mice (WT) in the water-maze. A) Average path length to reach the platform. Both group improved their performance similarly. B) Percentage of strategy used by L66$^{+/+}$ and WT mice on Day 1 and 4 analysed by means of a visual manual scoring or C) F-logic.

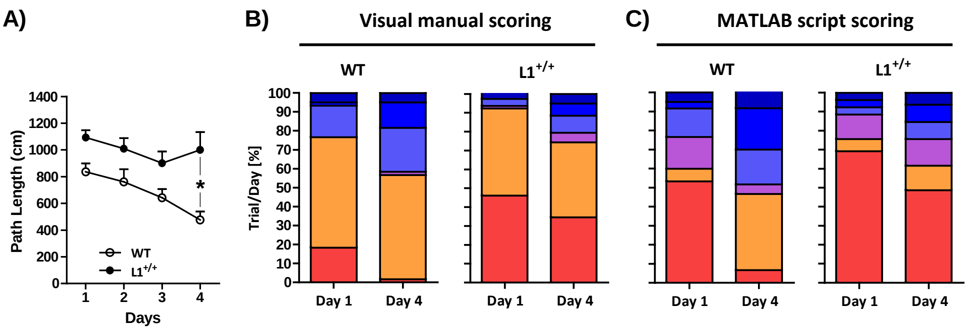


Figure 3. Comparison between visual and automatic analysis regarding the strategy used by transgenic mice (L1$^{+/+}$) and wild type mice (WT) in the water-maze. A) Average path length to reach the platform. Line 1 mice showed slower learning and differed significantly from WT mice on the 4th day. *P < 0.01, significantly different from control, Bonferroni test. B) Percentage of strategy used by L1$^{+/+}$ and WT mice on Day 1 and 4 analysed by means of a visual manual scoring or C) F-logic. Note the genotypic and categorisation-related differences.

References

[1] Morris, R. (1981). Spatial localisation does not depend on the presence of local cues. *Learning and Motivation*, ***12***, 239-260.

[2] Wolfer D, Lipp H. (2000). Dissecting the behaviour of transgenic mice: Is it the mutation, the genetic background, or the environment? *Experimental Physiology*, ***85***, 627-634.

[3] Garthe, A., Behr, J., Kempermann, G. (2009). Adult-Generated Hippocampal Neurons Allow the Flexible Use of Spatially Precise Learning Strategies. *Plos ONE*, ***4****(5)*.

Automated Analysis of Spatial Search Strategies in the Open-field Water Maze and Barnes Maze

B. Crouch and G. Riedel

School of Medical Science, University of Aberdeen, Aberdeen, Scotland, United Kingdom

Strategic analysis of the open field water maze [1] describes the process of classifying the swim path of an animal as it attempts to locate the submerged target platform. There is little consistency in the literature as to the number of strategies presumed to exist or the conditions which characterize each strategy. However the majority of established classification schemes [2,3] predict 3 main classes of search strategy, these are termed *Spatial*, *Procedural* and *Random*. Commonly used sub-divisions of these classes include *Direct*, *Focal* and *Directed* searches within the spatial class and *Chaining* or *Scanning* behaviour within the procedural class (see Figure 1). While search strategies of the open field water maze have been relatively well established there has been very little progress toward classifying the strategies associated with solving alternative tests of visuo-spatial working memory such as the Barnes maze [4] beyond basic segmentation into spatial, procedural and non-spatial groups.

While historically the classification of search paths has been carried out through a process of manual scoring, there have been motions in the literature toward the combination of multiple behavioural metrics into discriminant functions which enable the automated and unbiased scoring of search strategies. In order to see widespread adoption however, a classification system must meet several challenges, most notably the issue of flexibility. Systems must adapt to differences in maze parameters across labs (Pool size etc.) and to genotype differences within the same search pattern. There are also considerable costs (ethical, legal and monetary) associated with repeating a previous study making it impractical to carry out new experiments in order for the experimental protocol to comply with the requirements of the discriminant function. Advances in high density storage have enabled labs to archive large volumes of behavioural data which a sufficiently flexible analysis system could be used to analyse without need for further experimentation.

Presently the preferred method of automated strategy analysis is one of hierarchic sorting by testing multiple input parameters against arbitrary threshold values. If an individual path meets all criteria for a particular strategy it will be classified accordingly, if the track fails to meet the criteria it will then be disqualified and tested against the criteria for the next strategy down in the processing hierarchy. The weakness of such systems emerges when one strategy narrowly misses the threshold for classification and instead is falsely disqualified and wrongly assigned to another strategy. The first consequence of this flaw is that the strategies assessed further down the processing chain will be vulnerable to artificial dilation. The second is that minute changes in pool setup or animal behaviour can change the optimal threshold settings meaning that in a rigid system the criteria for classification can easily become incorrect disrupting the validity of the results. Here we present an alternative methodological approach to the use of “hard” thresholds and hierarchic sorting, substituting instead a highly flexible pseudo – fuzzy logic (see Figure 2) based sorting system. 1500 tracks from a pre-existing archive of animal behavioural data were manually scored and values for 11 different behavioural parameters were collected for each track. The tracks were then segregated by strategy and means / standard deviations of each parameter were calculated for each strategy. This data was then used to generate Gaussian membership functions which defined the expected input range for each parameter in each strategy. Rather than producing a binary pass or fail for any given input value these membership functions produce a degree of membership as their output. These outputs are then aggregated to produce a degree of confidence that a track belongs to a given strategy. This system allows for redundancy, meaning that a track can fail to meet selection criterion on a small number of individual parameters but still be correctly classified providing that its match on other parameters is sufficiently strong to counterbalance the error. This approach also allows the system to compensate for changes in pool parameters. Small changes are absorbed by the fuzzy sets while for large changes (such as cross-lab equipment differences) the Gaussian sets can easily be recalculated to match the experimenter’s setup using a small training set.

We will present a demonstration and comparison of 2 newly generated experimental software platforms which offer automated analysis of spatial search strategy. One of these systems is based on the conventional method of hard limiting. The other is based on the aforementioned fuzzy logic methodology. Each of these systems exists in two variants one optimized for the water maze and one for the Barnes maze.

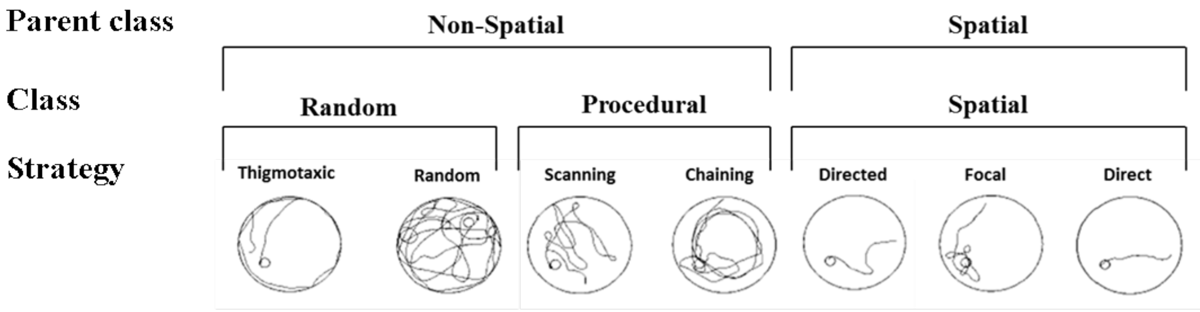


Figure 1. Generalization of a commonly used water-maze search strategy classification system. Parent classes are defined as spatial or non-spatial by the presence or absence of egocentric search behaviour respectively. Within the non-spatial parent class strategies are defined as either procedural or random. Finally individual strategies are defined by path-form. Direct = high efficiency linear search, Focal = focused search in close proximity to target, Directed = low efficiency axial search, Chaining = circular search path at fixed distance from pool wall, Scanning = search confined to pool centre, Random = high area coverage, no discernable strategy, Thigmotaxic = "wall hugging" behaviour. During learning of the spatial location of the platform, rodents typically migrate through these classes and become more accurate and spatial in their performance.

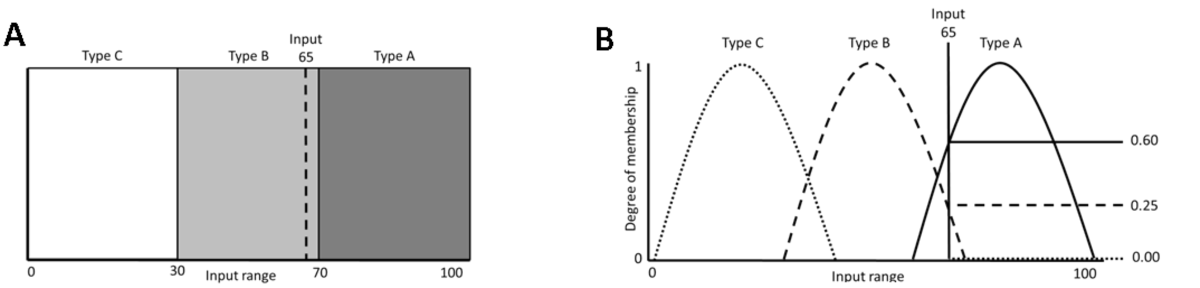


Figure 2. Crisp (binary) and fuzzy logic. In a conventional logic system (A) a strategy is defined by whether input values are above or below arbitrary threshold limits. In this example a classification of strategy Type A requires an input value above 70, Type B requires an input between 30 and 70 and Type C requires an input below 30. In this case a track in which all input parameters but one strongly indicate a Type A strategy is the best fit would still be disqualified from classification as Type A regardless of how small the distance from threshold is. In the fuzzy inference system (B) there is no clear point of distinction between Type A and Type B. instead there is a transitional zone where membership of multiple strategy types is possible. In example (B) when the input is 65 then the "degree of membership" of strategy Type A is 0.60 and Type B is 0.25. The strategy indicated is the one, which has the highest cumulative degree of membership for all input parameters.

References

[1] Morris, R. (1981). Spatial localisation does not depend on the presence of local cues. *Learning and Motivation*, ***12***, 239-260.

[2] Garthe, A., Behr, J., Kempermann, G. (2009). Adult-Generated Hippocampal Neurons Allow the Flexible Use of Spatially Precise Learning Strategies. *Plos one*, ***4****(5).*

[3] Janus, C. (2004). Search strategies used by APP transgenic mice during navigation in the Morris water maze. *Learning and Memory*, ***11****(3)*, 337-46.

[4] Barnes, C. (1979). Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J Comp Physiol Psychol*, ***93***, 74-104.

Non-genomic effects of testosterone on social behavior in laboratory rats

J. Hodosy[1,3], B. Filová[1], V. Borbélyová[1], J. Bábíčková[1], E. Domonkos[2] and P. Celec[1,4,5]

[1]Institute of Molecular Biomedicine, Comenius University, Bratislava, Slovakia.
[1]hodosy@gmail.com
[2]Department of Animal Physiology and Ethology, Comenius University, Bratislava, Slovakia
[3]Institute of Physiology, Comenius University, Bratislava, Slovakia
[4]Deparment of Molecular Biology, Comenius University, Bratislava, Slovakia
[5]Institute of Pathophysiology, Comenius University, Bratislava, Slovakia

Introduction

Testosterone is a steroid hormone, which has a crucial role for development of the brain and its functions, including behavior and cognitive abilities. It influences the development of reproductive system and secondary sexual characteristics, distribution of fat and hair growth [1]. In men, testosterone is mainly produced by Leydig cells of the testes [2], while in females it is mainly generated by the ovaries and adrenal cortex. Molecular mechanism of testosterone action can vary. Testosterone can act through slower genomic pathway or through rapid non-genomic pathway [3]. Studies that examine the non-genomic effect of testosterone on male and female social behavior are still rare. The aim of our work was to find out the non-genomic effect of testosterone on the social behavior of OVX (ovariectomized) female and GDX (gonadoectomized) male rats after the administration of androgen and estrogen receptor blocker to exclude the genomic pathway.

Experimental procedures and surgery

In our experiment, we used 20 female and 40 male Wistar rats (Anlab, Prague, Czech Republic). We housed the animals in separate polycarbonate cages in a controlled environment (temperature 22±2°C, humidity 50%±10%) with a 12:12 light – dark cycle, light phase from 08:00 a.m. to 8:00 p.m. The animals had *ad libitum* access to a standard diet and tap water. We randomly divided the female rats into 2 groups: control group (OVX-CTRL, n = 10) and testosterone group (OVX-TST, n = 10) and male rats into 4 groups: sham operated group treated with oil (SHAM-CTRL, n = 10), sham operated group with testosterone (SHAM-TST, n = 10), castrated group treated with oil (GDX-CTRL, n = 10) and castrated group with testosterone (GDX-TST, n = 10).

Male rats on postnatal day (PND) 47 underwent a castration (removal of testes) or sham surgery (without removal of the testes) under general anesthesia (intraperitoneal injection of ketamine (100mg/kg) and xylazine (10 mg/kg)). All female rats at age of 15 weeks underwent ovariectomy, when we extracted both ovaries. All animals were allowed two weeks for recovery after the surgery.

Hormonal supplementation and behavioral testing

We tested the rats in period of adulthood (PND100 – 125) in the test of social interaction. The apparatus consisted of dark plastic square arena 100cm x 100cm. In a corner of test arena we placed a cage with novel stranger rat of the same gender as testing animal. We put the tested animal in the middle of the maze and allowed to freely explore the maze for 10 minutes. The examined behaviors included: total distance moved, average velocity of the movement and time spent with investigating the stranger rat. The time spent with a novel stranger rat expressed the capability of social interaction and reflected the degree of exploratory and anxiety-like behavior. Rats received hormonal supplementation 1 hour before behavioral testing and group assignment was as follows: OVX-CTRL (olive oil – 1 μl/g); OVX-TST (testosterone propionate – 1 mg/kg); SHAM-CTRL (olive oil – 1 μl/g); SHAM-TST (testosterone propionate – 1 mg/kg), GDX-CTRL (olive oil – 1 μl/g); GDX-TST (testosterone propionate – 1 mg/kg). Animals received flutamide, an androgen receptor blocker (20 mg/kg) and tamoxifen, nonselective estrogen receptor blocker (10 mg/kg) one day before testing, and subsequently 1 hour before it. We collected blood samples from tail vein after behavioral testing and subsequently measured TST levels in plasma using a commercially available ELISA kit (DRG Diagnostic, Marburg, Germany). We analyzed the behavior of animals by a computerized animal observation system (EthoVision XT 10, Noldus, Netherlands).

Statistical analysis

We analyzed data from social interaction testing and the concentration of testosterone of males using the one-way analysis of variance (ANOVA) with the four groups being the tested factor. For comparison between pairs of groups we used Bonferroni post-hoc test. We analyzed data from behavioral testing and concentration of testosterone of females using unpaired nonparametric Mann-Whitney test. P-values less than 0.05 were considered significant. We used GraphPad Prism 5 for all calculations and tests. Data presented as mean + standard deviation.

Results

We found significantly increased concentration of testosterone in males in SHAM-CTRL ($F(3.36) = 12.11$, $p<0.01$), SHAM-TST ($p<0.001$) and GDX-TST group ($p<0.001$) in comparison with GDX-CTRL group. Testosterone levels did not differ between SHAM-CTRL, SHAM-TST and GDX-TST groups. OVX-TST group of females had significantly higher concentration of testosterone ($p<0.001$) when compared with OVX-CTRL group. In test of social interaction, we did not observe significant differences in time spent in interaction with the stranger rat between tested groups of males ($F(3.35) = 0.2300$) and females. Similarly, we found no significant differences in distance moved and velocity of the movement.

Conclusion

Our experiment did not confirm the rapid non-genomic effect of testosterone on social behavior after blockade of androgen and both estrogen receptors neither in male nor in female rats. Although, the concentration of testosterone in plasma was higher in groups received testosterone propionate, we did not find significant differences in social behavior of tested animals. Further studies are needed to uncover the molecular mechanism of non-genomic effect of testosterone on social behavior.

Ethical statement

All animal procedures were approved by the ethical committee of the Faculty of Medicine, Comenius University.

Acknowledgement

The study was supported by grants No. APVV-0753-10 and APVV-0539-12.

References

[1] Durdiaková, J., Ostatníková, D., Celec, P. (2011). Testosterone and its metabolites--modulators of brain functions. *Acta Neurobiologiae Expermientalis,* ***71***, 434-454.

[2] Schmidtová, E. (2008). Testosterone effects, metabolism and genetic determination. *Československá Fysiologie,* ***57***, 61-75.

[3] McEwen, B.S. (1991). Non-genomic and genomic effects of steroids on neural activity. *Trends in Pharmacological Sciences,* ***12***, 141-147.

Effects of increased prenatal testosterone on anxiety-like behavior and spatial memory of laboratory rats

V. Borbélyová[1], B. Konečná[1], E. Domonkos[2], J. Hodosy[1,3] and P. Celec[1,4,5]

[1]Institute of Molecular Biomedicine, Comenius University, Bratislava, Slovakia.
[1] borbelyova.veronika88@gmail.com
[2]Department of Animal Physiology and Ethology, Comenius University, Bratislava, Slovakia
[3]Institute of Physiology, Comenius University, Bratislava, Slovakia
[4]Deparment of Molecular Biology, Comenius University, Bratislava, Slovakia
[5]Institute of Pathophysiology, Comenius University, Bratislava, Slovakia

Introduction

Testosterone belongs to steroid hormones from the androgen group, which are important not only for sexual differentiation of the fetus, but also for the development of the nervous system and many aspects of behavior and cognitive abilities. Morphologically and physiologically testosterone is a determinant of male sex, it influences the development of reproductive system and secondary sexual characteristics, distribution of fat and hair growth [1]. Testosterone is mainly produced by the Leydig cells of the testes and fewer extent in kidney, liver, adrenal cortex [2] and in other peripheral tissues (fat and muscle tissue), even in some regions of the brain. Testosterone like other gonadal hormones have double effects: organizational and activational. Organizational (prenatal) effects of hormones occur during critical periods of development when exposure to gonadal hormones can cause permanent sex differences. Thus, organizational effects are permanent, irreversible and are responsible for sexual differentiation of the brain [3]. Later in life, testosterone has got activational effect on brain structures, which is dependent on its changing levels. Thus activational effects are acute and reversible that occur throughout life [4]. The aim of our study was to evaluate the organizational effects of testosterone on anxiety and spatial memory in juvenile and adult male and female rats.

Materials and methods

In this study, we used pregnant Lewis rats, which were kept in separate polycarbonate cages and were randomly divided into 4 groups: control (CTRL), testosterone propionate (TST), flutamide (FLU), testosterone propionate with flutamide (TST+FLU). From 14th day of pregnancy until delivery, the rats received either testosterone propionate in dose of 2mg/kg, flutamide in dose of 5mg/kg, both testosterone propionate and flutamide, or olive oil intramuscularly. We housed pups after delivery with their mothers until weaning on postnatal day (PND) 21 and we divided young rats into cages according to sex and birth nest. The housing room was temperature controlled (temperature 22°C, humidity 50%) with a 12:12 light – dark cycle, light phase from 08:00 a.m. to 8:00 p.m. The animals had *ad libitum* access to a standard diet and tap water.

We tested the offsprings in several behavioral tests in pre-adolescent period (PND 25 – 27) and in adulthood (PND 82 – 124). Behavioral testing was conducted during the light phase, although in dark room. In pre-adolescent period, we investigated the behavioral performance of rats in open field test, light/dark box and elevated plus-maze to assess anxiety-like behavior and in simple novelty test to evaluate their memory. We tested rats in period of adulthood in above mentioned tests and in an additional test, in modified Morris water maze to assess their spatial memory.

The **open field apparatus** (5 min) consisted of dark plastic square arena of 100cm x 100cm. The examined behavior included: total distance moved, average velocity of the movement and time spent in the centre zone. The **light/dark apparatus** (5 min) consisted of dark plastic rectangle arena of 45cm x 65cm. It consisted of two equally sized chambers connected by an opening. One chamber was black and covered with a black lid. The second chamber was opened and brightly illuminated. The examined behavior was the amount of time spent in the light chamber. The **plus-maze apparatus** (5 min) consisted of dark plastic arena with two opened arms and two enclosed arms by walls (30cm high). The arms extended from a central platform. The maze was connected to a metal frame on each end of the arms raising it 60cm above the floor. We recorded the amount of time spent in opened arms. **Simple novelty test** (5 min) we realized in the open field apparatus, but we placed a novel object to the test arena. The examined behaviors included: total distance moved, average velocity of the movement and time spent with investigating a novel object. The **modified Morris water maze** test we realized in a circular pool (diameter 125cm, height 60cm) filled with tap water (24-26°C), which we placed in a brightly illuminated room. The maze was virtually divided into four quadrants and one geometrical figure for orientation as an external cue was placed on the wall of each quadrant. We placed an escape platform (diameter 10cm) into one of four quadrants, hidden 2 cm below the surface of the water (the platform position remained the same during the testing period). The animals were tested in blocks of four trials (a trial duration – 60s) a day, for 4 consecutive days (a same order of animals and starting quadrant at each trial for 4 consecutive days). A trial (60s) consisted of a swim followed by a 30s rest on the platform. If the rat did not find the platform within 60s, the investigator navigated it to the platform. The examined behavior was the time to find the platform. On the day 5 a probe trial occurred. We removed the platform during this trial from the pool. The duration of this trial was 60s and time spent in the platform quadrant was recorded (long – term or reference memory). The starting position for all animals was the quadrant opposite to the platform quadrant.

We measured the behavior of the animals via a computerized animal observation system (EthoVision XT 10, Noldus, Netherlands). We collected blood samples from pregnant females daily, 1 h after injection, from 18th day of pregnancy until the delivery and subsequently measured TST levels in plasma using a commercially available ELISA kit (DRG Diagnostic, Marburg, Germany). We analyzed data using ANOVA with Bonferroni post-hoc test. $P<0.05$ was considered to be statistically significant. We used GraphPad Prism 5.0 for all calculations and tests. Data are presented as mean + standard deviation.

Results

We found significantly higher testosterone levels in plasma of pregnant females in the TST and TST+FLU groups when compared to the CTRL group ($p<0.0001$), while testosterone levels did not differ between CTRL and FLU groups.

Experimental groups differed neither in the analyzed behavioral parameters in any of behavioral tests (open field, elevated plus maze, light/dark box, simple novelty test) nor in preadolescent period and adulthood. In the water maze (period of adulthood), we detected general improvement of animals within acquisition trials. All animals showed improvement from day 1 to day 4, however we observed significant difference between TST and FLU groups on day 4 only (*p<0,05). During the probe trial, all groups spent equal time in the quadrant where the platform was previously hidden.

Conclusion

In our experiment prenatal administration of testosterone had no effect on anxiety behavior neither in the preadolescent period nor in period of adulthood. The effects of the testosterone on spatial reference memory remains less clear. In Morris water maze we found significant effect of prenatal testosterone only on 4th day of the test, when TST group was better in finding the platform than FLU group. Studies about the organizational effect of testosterone on behavior are still rare and their conclusions are often contradictory and more research could be aimed on searching strategies during water maze task.

Ethical statement

All animal procedures were approved by the ethical committee of the Faculty of Medicine, Comenius University.

Acknowledgement

The study was supported by grants No. APVV-0753-10 and APVV-0539-12.

References

[1] Durdiaková, J., Ostatníková, D., Celec, P. (2011). Testosterone and its metabolites--modulators of brain functions. *Acta Neurobiologiae Expermientalis,* ***71***, 434-454.

[2] Schmidtova, E. (2008). Testosterone effects, metabolism and genetic determination. *Československá Fysiologie* , ***57***, 61-75.

[3] Morris, J.A., Jordan, C.L., Breedlove, S.M. (2004). Sexual differentiation of the vertebrate nervous system. *Nature Neuroscience* , ***7***, 1034-1039.

[4] Williams, C.L. (1986). A reevaluation of the concept of separable periods of organizational and activational actions of estrogens in development of brain and behavior. *Annals of the New York Academy of Sciences,* ***474***, 282-292.

Prior test experience produces changes of t-patterns spatial distribution in the Elevated Plus Maze test

M. Casarrubea[1](*), A. Santangelo[1], F. Faulisi[1], V. Roy[3], A. Arabo[3], F. Sorbera[1], M.S. Magnusson[2], G. Crescimanno[1]

[1] Dept. of Experimental Biomedicine and Clinical Neurosciences (BioNec), Human Physiology Section "G. Pagano", Laboratory of Behavioral Physiology, University of Palermo - Palermo, Italy

[2] Human Behavior Laboratory, University of Iceland - Reykjavik, Iceland

[3] PSY-NCA, EA4700, Laboratoire de Psychologie et de Neurosciences de la Cognition et de l'Affectivité, Université de Rouen, Mont-Saint-Aignan, France

(*) Corresponding Author: maurizio.casarrubea@unipa.it

Abstract

Aim of present research was to investigate in male Wistar rats whether a prior elevated plus maze experience modifies the temporal structure of the behavioral response following a retest applied after 24h. Video files were coded by means of a software coder and event log files generated for each subject were analyzed by means of a specific software for temporal pattern analysis (Theme). Present research shows a clear reduction of the number of t-patterns from trial one to trial two. This reduction is provoked by the disappearance of t-patterns consisting of behavioral elements occurring in the unprotected zones of the maze. The results suggest that the previous experience in the maze causes learning-dependent behavioral changes inducing a more clear-cut response to environmental anxiogenic conditions.

Introduction

The elevated plus maze (EPM) is a behavioral model widely used to assess anxiety-related behavior in rodents. The utilization of EPMs is based on the premise that the simultaneous presence of the enclosed arms (scantly illuminated and surrounded by walls) and of the open arms (brightly illuminated and without walls) evokes a typical approach/avoidance conflict [2][9]. Today, the EPM is also one of the most commonly used assays to investigate the biological basis of emotionality. In addition, thanks to the test/retest protocol, researchers have extended its employment to the study of learning and memory [5][8]. By means of a multivariate approach, known as t-pattern analysis [7], we have already demonstrated the existence of a behavioral temporal structure characterizing the anxiety-related activity of Wistar rats tested in EPM [3]. Aim of present research was to determine in Wistar rats tested in EPM whether a prior maze experience was able to induce meaningful changes in the temporal structure of behavior analyzed following a re-test applied after 24h.

Method

Twenty, three months old, pathogen free, male Wistar rats were used. Animals were born in the animal facility of the University of Rouen (France) and breeders originated from Janvier (Le Genest-St-Isle, France). Rats were housed in groups of three in a room maintained at the constant temperature of 21 ± 2 °C, under the following light/dark cycle: light on = 12 noon; light off = 12 midnight. Food and water were available ad libitum. Each rat, experimentally naïve, was exposed to EPM for two times according to a test/retest protocol [2]. Retest was scheduled after an interval of 24 hours from the prior test. Each subject was placed in the central platform of EPM and allowed to freely explore for 5 min [10]. Experiments were recorded through a video camera and video files stored in a personal computer for following analyses. In present study we have employed the ethogram described in Table 1. As a result of the coding process, an event log file was obtained. This is a sequence of behavioral events chronologically ordered on the basis of their onset time. In this study, each video was coded following a frame-by-frame examination using the software The Observer (Noldus Information Technology, The Netherlands). To assess the temporal relationships among behavioral events, log files were processed by means of the software program Theme (Noldus Information Technology bv, The Netherlands; Patternvision Ltd, Iceland).

Table 1. Ethogram of rat behavior in the elevated plus maze test. (*) = the behavioral element is considered protected (p-) when performed in the central platform or in a closed arm, unprotected (u-) when performed in an open arm; (**) = head dip is considered protected (p-) only in the central platform and unprotected (u-) in an open arm.

Behavioral element	Abbreviation	Description
Closed Arm Entry	CA-Ent	rat moves from the Central Platform to a Closed arm (all four paws in)
Open Arm Entry	OA-Ent	rat moves from the Central Platform to an Open Arm (all four paws in)
Closed Arm Return	CA-Ret	rat puts only head and forepaws in the central platform, then rapidly re-enters in the closed
Closed Arm Walk	CA-Wa	rat walks in a Closed Arm
Open Arm Walk	OA-Wa	rat walks in an Open Arm
Central Platform Entry	CP-Ent	rat moves from an Open or Closed Arm to the Central Platform
Immobile Sniffing (*)	p-ISn; u-ISn	rat sniffs the surrounding area without walking activity
Corner Sniffing (*)	p-CSn; u-CSn	rat sniffs the entrance border of a Closed Arm
Stretched Attend Posture (*)	p-SAP; u-SAP	rat stretches its head and shoulders forward and then returns to the original position
Head Dip (**)	p-HDip; u-HDip	scanning over the sides of the maze towards the floor
Rearing (*)	p-Re; u-Re	rat maintains an erect posture
Defecation (*)	p-Def; u-Def	excrements are produced
Grooming (*)	p-Gr; u-Gr	rat licks/rubs its face and/or body
Paw Licking (*)	p-PL; u-PL	rat licks its paws
Immobility (*)	p-Imm; u-Imm	an immobile posture is maintained

Results

113 t-patterns of different composition were detected in trial-1, only 24 in the retest (Figure 1). Of the 113 t-patterns, 61 consisted of behavioral elements carried out in open arms and central platform (OA-CP), 28 consisted exclusively of behavioral elements carried out in the closed arms and central platform (CA-CP), and the remaining 24 showed a mixed composition (OA-CP and CA-CP). The t-patterns of trial-1 were composed by a total amount of 11 elements of which 6 protected (CA-Ent, Ca-Wa, CP-Ent, p-Csn, p-Isn, p-Re) and 5 unprotected (OA-Ent, OA-Wa, u-Csn, u-HDip, u-Isn). In trial-2, the 24 t-patterns observed were composed only by 6 behavioral elements carried out in the protected zones of the EPM (CA-Ent, Ca-Wa, CP-Ent, p-Csn, p-Isn, p-Re). Finally, the 113 different t-patterns of trial-1 occurred (mean ± SE) 482.65 ± 24.85 times (Figure 2); on the other hand the 24 different t-patterns of trial-2 occurred 227.6 ± 13.79 (Figure 2). Student's *t*-test revealed a highly significant ($p < 0.0001$) difference between these mean values.

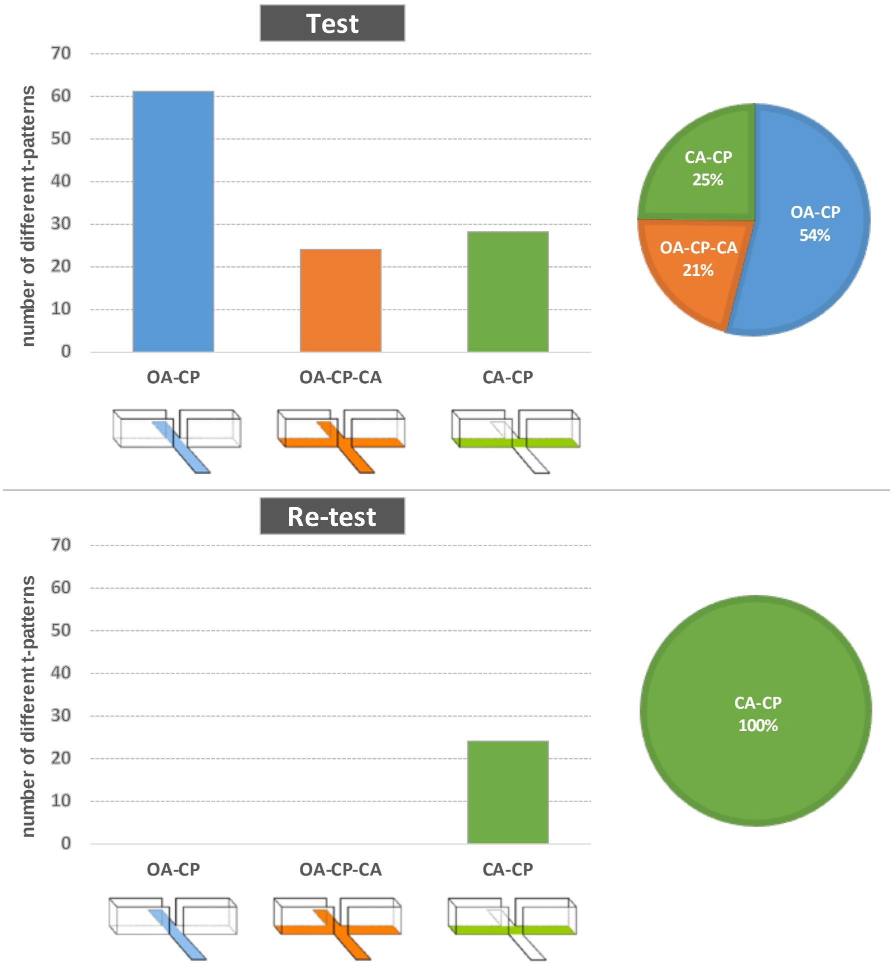


Figure 1. Number of different t-patterns and percent distributions (pie charts) in OA-CP, OA-CP-CA and in CA-CP.

Discussion

Present data show that in the Wistar rat tested in EPM the number of t-patterns undergoes a clear reduction from trial-1 to trial-2 (Figure 1 and Figure 2). This reduction is represented by the disappearance of t-patterns carried out in OA-CP and in OA-CP-CA (Figure 1). A new environment such as the EPM provokes a strong stress, inducing an acute fear condition. Fear triggers an active behavioral strategy that, as proposed by Walter Cannon [1], is aimed at removing the source of danger through the escape. During test one, when the impossibility of the escape is ascertained, escape attempts are progressively abandoned. In the second half of test one, when the fear changes into anxiety and the animal has collected the information on the risk, a strong approach-avoidance conflict occurs. The conflict is solved through the choice of the closed arms where the animal finds refuge and shows an activity characterized by t-patterns consisting of exploratory behavioral elements (CA-Ent, CA-Wa, CP-Ent; p-Csn, p-Isn, p-Re). In the retest, as primary purpose, the rat moves toward the protected zones (CA and CP) of the EPM, while a clear aversion is displayed toward the unprotected ones (OA). Indeed, the number of t-patterns decreases (Figure 2) and they consist of behavioral elements occurring only in CA and CP (Figure 1). These data can be interpreted as the effort of the animal to organize behavioral patterns aimed at a more prompt solution of the approach-avoidance conflict. In conclusion, according to the literature [2][4][6] our results suggest that the previous experience in the maze causes learning-dependent behavioral changes inducing a more clear-cut response to environmental anxiogenic conditions.

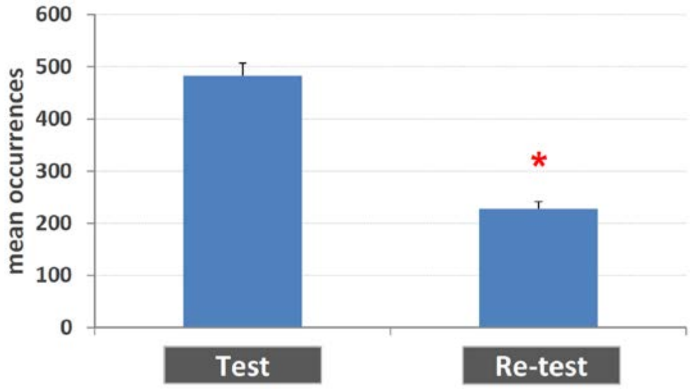


Figure 2. Mean number of t-patterns for each subject. * significant ($p < 0.0001$) difference (Student's *t*-test)

Ethical Statement

All efforts were made to minimize the number of animals used and their suffering. All the experimental procedures were conducted in accordance with the European Communities Council Directive (86/609/EEC) and approved by the Veterinary Committee officially appointed by the University of Palermo.

References

[1] Cannon, W.B. (1915). *Bodily changes in pain, hunger, fear and rage*. New York, NY; Appleton.

[2] Carobrez, A.P., Bertoglio, L.J. (2005). Ethological and temporal analyses of anxiety-like behavior: The elevated plus-maze model 20 years on. *Neuroscience & Biobehavioural Review*, ***29***, 1193-1205.

[3] Casarrubea, M., Roy, V., Sorbera, F., Magnusson, M.S., Santangelo, A., Arabo, A., Crescimanno, G. (2013). Temporal structure of the rat's behavior in elevated plus maze test. *Behavioural Brain Research*, ***237***, 290-299.

[4] Cruz-Morales, S.E., Santos, N.R., Brandao, M.L. (2002). One trial tolerance to midazolam is due to enhancement of fear and reduction of anxiolytic-sensitive behaviors in the elevated plus maze retest in the rat. *Pharmacology Biochemistry & Behavior*, ***72***, 973-978.

[5] Gazarini, L., Stern, C.A., Bertoglio, L.J. (2011). Protein synthesis in dorsal hippocampus supports the drug tolerance induced by prior elevated plus-maze experience. *Neuroscience*, ***179***,179-187.

[6] Holmes, A., Rodgers, R.J. (1998). Responses of Swiss–Webster Mice to Repeated Plus-Maze Experience: Further Evidence for a Qualitative Shift in Emotional State? *Pharmacology Biochemistry & Behavior*, ***60***,473-88.

[7] Magnusson, M.S. (2000). Discovering hidden time patterns in behavior: t-patterns and their detection. *Behavioural Research Methods Instruments & Computer*, ***32***, 93-110.

[8] Rodgers, R.J., Johnson, N.J., Cole, J.C., Dewar, C.V., Kidd, G.R., Kimpson, P.H. (1996). Plus-maze retest profile in mice: importance of initial stages of trail 1 and response to post-trail cholinergic receptor blockade. *Pharmacology Biochemistry & Behavior*, ***54***, 41-50.

[9] Roy, V., Chapillon, P., Jeljeli, M., Caston, J., Belzung, C. (2009). Free versus forced exposure to an elevated plus maze: evidence for new behavioral interpretation during test and retest. *Psychopharmacology*, ***203***, 131-141.

[10] Walf, A.A., Frye, C.A. (2007). The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nature Protocols*, ***2***, 322-328.

Thermal response behaviors in larval zebrafish: startle escape, thermotaxis and thermal arousal

Tohei Yokogawa[1], Michael Iadarola[2] and Harold Burgess[1]

[1]NIH-National Institute of Child Health and Human Development
[2]NIH-Clinical Center
yokogawat@mail.nih.gov

Detection of the changing thermal environment is critical for animal survival. Sensory detection drives motor behaviors which allow animals to rapidly escape from noxious temperatures and initiate search patterns to move back into an optimal thermal environment.
Larval zebrafish have multiple advantages for examining motor behavior and understanding neural circuits. These include: small size for simultaneously recording large numbers of fish, stereotypic swim patterns which can be computationally categorized [1], and the availability of transgenic fish for ablation, monitoring, tracing and activation of targeted neurons. However, very few studies have examined the response of zebrafish to thermal cues. Here we report three types of behavioral response to changes of the thermal environment in larval zebrafish.

Freely swimming zebrafish responses were monitored under three conditions: to a sudden high intensity thermal pulse, to a slow increase in ambient water temperature and to a temperature gradient across the testing arena. To probe the response of larvae to a sudden thermal stimulus, we aimed an infra-red laser at a small region within the testing arena which was triggered when a larva swam into the target zone. Larvae responded robustly to the acute high intensity thermal pulse, and kinematic analysis demonstrated that they performed a swim pattern similar to a Mauthner cell mediated escape response. In contrast, larvae in a flow chamber which were exposed to temperature gradient, showed thermotaxis, turning and swimming away from the region of high temperature. The temperature gradient rose from ambient temperature on one side to 5 C higher on the other side. Under uniform temperature conditions, larvae were evenly distributed across the testing arena. However after establishment of the temperature gradient, larvae aggregated on the room temperature side of the chamber. Kinematic analysis of swimming movements revealed that avoidance of the high temperature zone was due to three factors. First, larvae which were oriented toward the ambient zone showed a high frequency of forward swim movements driving them into this region. Second, fish within the high temperature zone showed a high frequency of turn initiations allowing a local search of water temperatures. Third, turns executed in the high temperature zone showed a strong directional bias, causing larvae to orient away from the high temperature area allowing them to then initiate forward swims toward the ambient zone. The third response to changes in environmental temperatures, thermal arousal, was manifest as a short term change in internal state. When ambient water temperature was gradually increased from baseline 26 C to a noxious 36 C, larvae became hyperactive. Importantly, hyperactivity persisted for several minutes after the water temperature returned to 26 C. We previously reported that water flow induces an arousal state [2], but under the conditions used in these experiments, we used slow flow rates to avoid flow-induced arousal. During thermal arousal, fish showed normal movement initiation frequency but increased velocity during swim bouts. We speculate that thermal arousal prepares fish to anticipate and quickly respond to subsequent changes in environmental temperature.

To investigate the neuronal basis for these different responses, we generated transgenic fish for optogenetic activation of individual trigeminal sensory neurons and tracing of axonal projections. For this, we took advantage of an enhancer trap line which strongly expresses Gal4 in a subset of trigeminal neurons. Ablation of trigeminal neurons using UAS:nitroreductase resulted in a reduction of thermal responses confirming the involvement of the neurons labeled in this line in these behaviors. By injecting a UAS:channelrhodopsin2 plasmid into this fish, we could activate trigeminal neurons in freely swimming fish using a pulse of intense blue light. We calibrated the plasmid dose so that individual fish stochastically expressed ChR2 in only a small subset of trigeminal neurons. ChR2 expressing fish initiated either escape responses or thermal arousal after trigeminal activation, demonstrating that distinct trigeminal neurons mediate these behaviors. Thus individual trigeminal neurons selectively drive different thermal behaviors. We then traced axons from trigeminal neurons with defined behavioral roles revealing distinct projection patterns. We are now using GCaMP imaging to identify neurons throughout the brain which respond to thermal stimuli. These studies will reveal key brain areas which mediate distinct modes of behavioral response to thermal cues.

Ethical statement

All in vivo experimental protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of the Division of Intramural Research of the National Institute of Child Health and Human Development. The NICHD Animal Care and Use program is administered by the Research Animal Management Branch (RAMB) which is staffed by board-certified laboratory animal veterinarians and pathologists, veterinary and animal care technicians.

References

[1] Fero, K., Yokogawa, T., Burgess, HA. (2010). The Behavioral Repertoire of Larval Zebrafish. In A.V. Kalueff and J.M. Cachat (Eds.), *Zebrafish Models in Neurobehavioral Research*. Humana Press.

[2] Yokogawa, T., Hannan, M., Burgess, HA. (2012). The dorsal raphe modulates sensory responsiveness during arousal in zebrafish. *Journal of Neuroscience, **32(43)***, 15205-15.

Determining stock cube uses in a kitchen - A naturalistic approach: Behaviors during the culinary preparation with different ingredients

C. Iborra-Bernad, E. Petit, A. Giboreau

Center for Food and Hospitality Research, Institut Paul Bocuse.
agnes.giboreau@institutpaulbocuse.com

Keywords: stock cubes, usage, observational methodology and experimental kitchen.

Introduction

In France a broth prepared over a long period of timewas a key element of traditional cooking [1]. In addition, the soup has been gastronomic dish since the XVIII century. The word “potage” instead of “soupe” (soup in English) was used to define the delicious soup made for the high society [2]. Around 1850, a chemist called Justus von Leibig (1803-1873), invented a technique to produce a meat extract on a large scale [2]. The use of stock cubes to flavour meals spread from the beginning of their industrial production in 1886, when Carl Heinrich Knorr started its commercialization [2]. Nowadays, stock cubes are commonly used in Europe but also in countries such as South Africa, as delivers taste and micronutrients without producing major changes in food production or changes in traditional diets [3].

In the present study, we rely on an ethnographic approach reproducing immersion on context of traditional ethnography in order to provide a direct experience and understanding of the users’ world for improve the design of the objects [4]. Thus, Salvador *et al.* [5] describe ethnography in the corporate context as "a way of understanding the particulars of daily life in such a way as to increase the success probability of a new product or service or, more appropriately, to reduce the probability of failure specifically due to a lack of understanding of the basic behaviours and frameworks of consumers."

In this research we are interested in learning in details how French consumers use stock cubes for cooking. Between different possible research methodologies, such as surveys, the observation in an experimental kitchen has been selected in this study. The observational measurements allowed taking into account the conscious and unconscious parts of behaviours [6], such as the different ways of using the stock cubes and the quantity used (data not showed).

Methods

We used a video recording which permit to capture complexity of real-life. In this sense, the complexity of the behaviours during cooking are coded combining the different options of the software The Observer XT (Noldus).

11 volunteers (average age of 48 years) habitual users of stock cubes participated in the study. Before starting, an ethical committee approved the study. A consent form for participation and video recording in the experimental kitchen was signed. Each cooking session observed lasted around 1h 30.

The study consisted of two main phases depending on the types of recipe. During phase 1, water, rice and stock cubes were available to prepare flavoured rice. During phase 2, water, rice, stock cubes, chicken, onion and sunflower oil were available to prepare fried rice. Ingredients given to conduct the tests were products from the same batch.

Three cameras located on the ceiling of the experimental kitchen were used to observe participants during the meal preparation. The record videos were employed to analyse participants’ behaviours towards the stock cubes. The Observer XT was employed to code the observed behaviours. The following information was checked: 1) What is the type of manipulation of the stock cubes before use?; 2) What are the ingredients flavoured with the stock cube?

Results

Observations led us to define various categories in answer to the previous questions. Figure 1 shows the type of manipulations and examples from the users. The main uses are adding the cube without handling and crumbling into the water, other ingredient or a mix of other ingredients. The stock cube was employed to season raw (chicken) or cooked products.

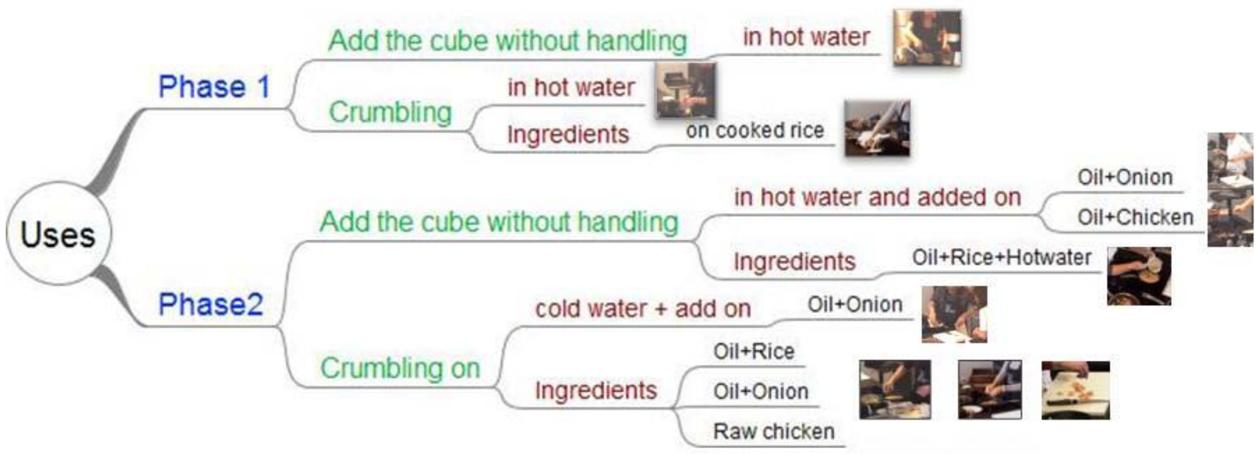


Figure 1. Diagram of uses according to the phase of the study.

Conclusions

In this work, we showed uses of stock cubes in different recipes with in the experimental kitchen. Stock cubes are used differently during the meal preparation for various purposes such as flavouring the broth. It was observed that adding the cube without handling is not the only use possible in meal preparation. Crumbling on ingredients like salt or dissolving the stock cubes in hot water to produce a concentrated sauce are relevant uses not expected before the study. The physico-chemical properties of the stock cubes will make the uses observed in our study easier, to satisfy the consumers' expectancies. This pattern of uses will be considered as the basis to develop a more complex study on the uses regarding traditional meals and ingredients of a region and culture.

The study could have limitations because the participants are not in their own kitchen. However, the experimental kitchen allows us to create standard situations and lower experimental variability by giving participants identical ingredients, recipes and equipment. In addition, this naturalistic setting could be an interesting context to conduct tests of products comparison, consumer knowledge and the assimilation of the innovative products by the potential users of a product.

References

[1] Girard-Lagorce, S. (2005). *Grandes et petites histoires de la gourmandise française: traditions et recettes*.Volume 5 de Terre de poche. Editions de Borée.

[2] Vander Sande, T. (2008).*Délices de bouillon*. Editions Racine Lannoo.

[3] Zhengxing, C., Wilna, O.-T. (2004).Household consumption of stock cubes and stock powder in the Vaal Triangle of SA. *Nutr. Food Sci., 34*, 174–178.

[4] Hanington B., Martin B. (2012). *Universal methods ofdDesign: 100 Ways to Research Complex Problems, develop innovative ideas, and design effective solutions.* Rockport Publishers.

[5] Salvador T, Bell G., Anderson K. (1999). Design ethnography. *Design management journal (former series), 10*, 35-41.

[6] Wansink, B., Sobal, J. (2007). *Mindless eating. environment and behavior*, *39*, 106-123.

Validation of a behavioral task to study perceptual supplementation in rats

V. Roy[1], O. Menant[1], R. Hacquemand[1], K. Rovira[1] and O. Gapenne[2]

[1] EA 4700, PSY-NCA, Université de Rouen, France.
vincent.roy@univ-rouen.fr
[2] CNRS-UMR 7338, BMBI, Université de Technologie de Compiègne, France.
olivier.gapenne@utc.fr

Abstract

This research validates a rodent model for perceptual supplementation (or sensory substitution). DA/Han rats were video-tracked in the darkness and had to find a reward by searching an auditory virtual track. Through the 3 stages of the experiment, rats learned to use the auditory feedback to follow the virtual track and reach the reward. Our results indicate that perceptual supplementation is possible in rats and our task could be adapted in order to study the mechanisms of perceptual supplementation and its ontogeny and neurobiological substrates.

Keywords: sensory substitution, perceptual supplementation, rats, behavior, learning

Introduction

Perceptual supplementation, also referred to as sensory substitution, is the process by which an individual can experiment its environment through unexpected sensations. For instance, the Tactile Vision Sensory System (TVSS) developed by Paul Bach-y-Rita allows blind people to perceive their outer world by manipulating a camera which, in return, sends tactile stimulations to the skin through a computerized device (Bach-y-Rita, 1969). Similarly, TACTOS was designed to study visuo-tactile exploration of a virtual environment (Ziat et al., 2007). By using TACTOS, a blind person is able to perceive shapes by moving a stylus on a graphics tablet and by receiving a tactile stimulation through a Braille cell every time the stylus runs over the line of a shape. The aim of our study was to develop a comparable situation of perceptual supplementation in rats. In this research, rats were video-tracked in darkness and they had to learn how to reach a reward by relying on an auditory signal that was played each time their snout entered a virtual track guiding to the reward. The present paper reports the protocol as well as the results that evidence for the first time the possibility of a perceptual supplementation process in rats.

Methods

DA/Han males rats (n=6) were tested from the age of 8 weeks. Animals were maintained in groups of 3 in standard cages with free access to food *ad libitum* (lights from 00:00 to 12:00). Behavioral testing took place from 13:00 to 16:00 and water access was restricted to 2 hours per day after the behavioral testing. Rats were tested in a 1m diameter semi-circular arena (Figure 1). The experimental room was totally dark and an ANY-maze® video-tracking system (Stoelting Co, Wood Dale, IL – version 4.89) tracked their movements under infrared lightening (3 projectors with 15 LEDs each). Each time the snout of the animal entered the correct virtual track among 3 possible, a speaker controlled by the ANY-maze interface played a continuous sound of 4 kHz and 55 dB. The sound by itself carried no spatial information on the reward location. When the snout of the animal reached the end of the track, a water reward was delivered at the floor level through a steel canula. The next trial started automatically when the animal reached back the intersection of the three possible virtual tracks.

The experiment was composed of 3 stages (see Figure 1). On each day of each stage, there were 40 trials given in 2 sessions of 20 trials with an intersession of approximately 30 minutes. For each trial, the correct track was randomly chosen by ANY-maze.

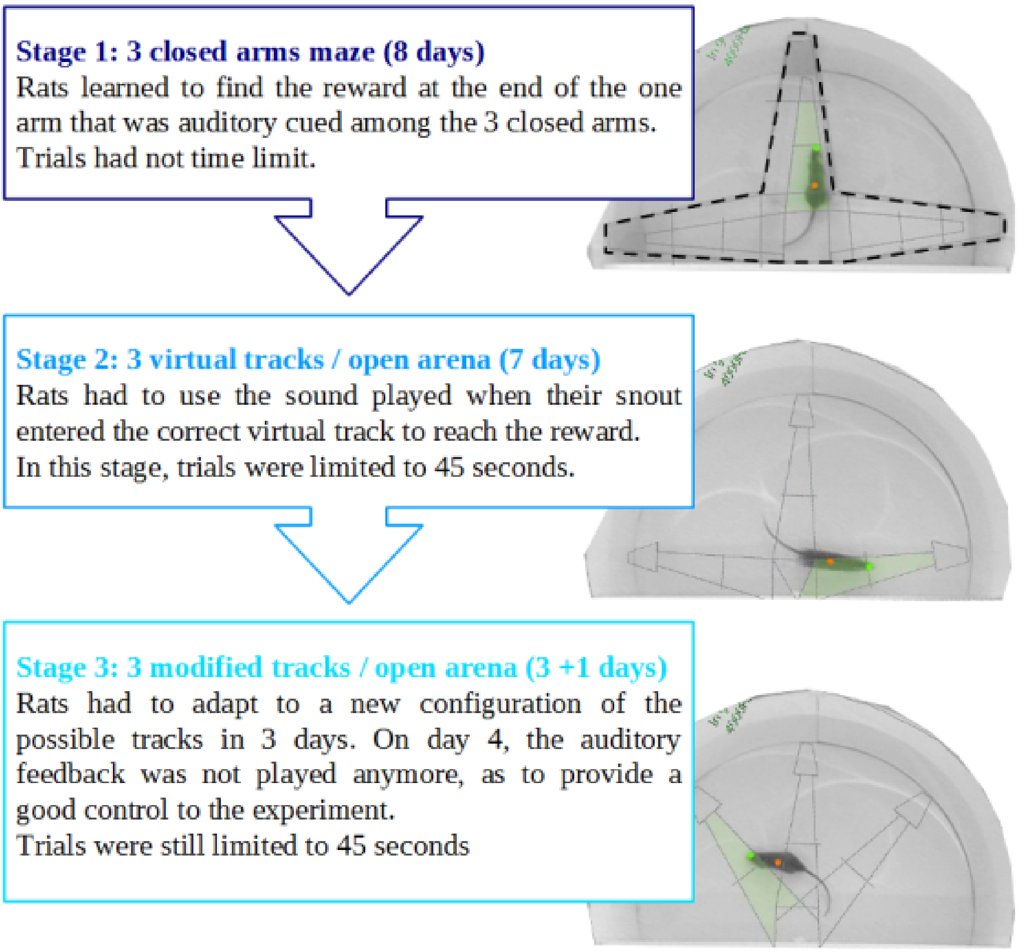


Figure 1. Illustration of the apparatus, tracking and schedule of the behavioral protocol. The 3 stages ran over 19 consecutive days. Dotted lines on the 1st stage indicate the position of the closed arms.

Stage 1 lasted 8 days and was conducted in a 3 closed arms maze. For each trial, rats learned to find the reward at the end of the arm that was auditory cued. Trials had no limit of time.

Stage 2 was 7 days long and took place in the semi-circular arena with 3 possible virtual tracks (same positions than the closed arms on stage 1). For each trial, rats had to use the sound played when their snout entered the correct virtual track to reach the reward. In order to limit alternate strategies and force the animal to use the auditory feedback, the reward was delivered only if the rat had entered both the proximal and the distal part of the track. Trial duration in stage 2 was limited to 45 seconds.

Stage 3 lasted 4 days. On the 3 first days, rats had to learn a new configuration of the tracks since the left and right ones were moved to an oblique position. On day 4, the auditory feedback was not played anymore in order to provide a good control to the experiment. Trials were still limited to 45 seconds.

Ethical statement: The research reported in this paper was conducted in accordance with the European council directive for the protection of animals used for experimental and other scientific purposes (EC86/609).

Results

Rats first learned to use the sound in the 3 arms maze. The number of errors, indicated by the number of time they went to the end of an arm not rewarded and not cued by the auditory feedback, decreased significantly in day 6, 7 and 8 compared to Day 1. On stage 2 (See Figure 2), the mean percentage of successful trials increased over days to reach a performance of 92.1% on day 7. This percentage was significantly more important than on day 1, 2 and 3. When the positions of two lateral tracks were changed on stage 3 (See Figure 3), the rats' performance dropped to 40% but reached 87.5% of success on day 3. On day 4 of stage 3, when the auditory feedback was removed, the rats' performance dropped to 47.5% (See Figure 3, white histogram).

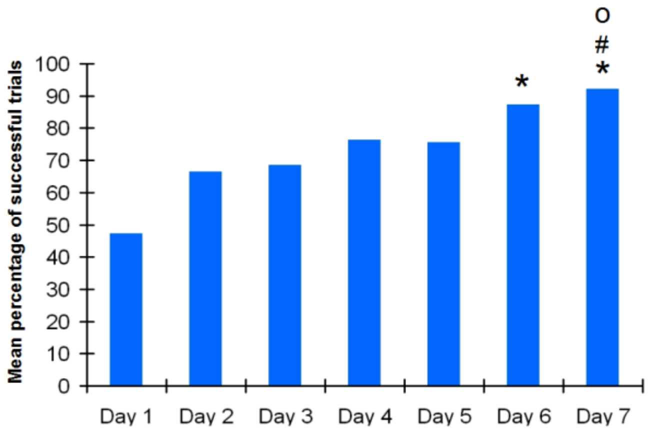


Figure 2. Mean percentage of success during stage 2. * $p<0.05$ in comparison to day 1; # $p<0.05$ in comparison to day 2; o $p<0.05$ in comparison to day 3.

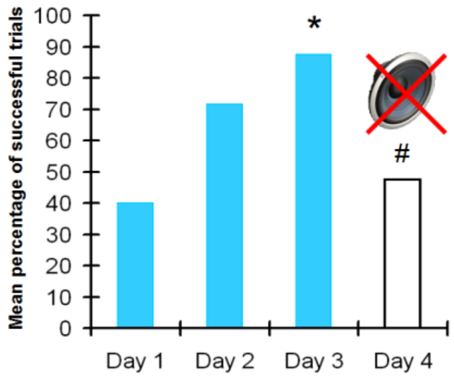


Figure 3. Mean percentage of success during stage 3. * $p<0.05$ in comparison to day 1; # $p<0.05$ in comparison to the day 3.

Discussion

Our research describes a valid protocol to study perceptual supplementation in rats. Animals learned to use efficiently the sound feedback to find the reward in all three stages of the experiment and, on the last day of stage 3, the percentage of success dropped drastically when the auditory feedback was removed. However, since 47% of the trials were still successful, it is probable that other information, such as idiothetic spatial information, were also used in this form of perceptual supplementation. In conclusion, our behavioral task may prove its importance in order to study the conditions that lead to perceptual supplementation as well as the ontogeny and the neurobiological substrates of this process.

Measuring human movement patterns and behaviors in public spaces: A method based on thermal cameras, computer vision, and geographical information system technologies

S. Z. Nielsen[1], R. Gade[2], T.B. Moeslund[3], H. Skov-Petersen[4]

[1]Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark.
[1] szn@ign.ku.dk, [4] hsp@ign.ku.dk
Visual Analysis of People Lab, Aalborg University, Aalborg, Denmark.
[3] rg@create.aau.dk, [4] tbm@create.aau.dk

In order to assess human movement patterns and behaviors in public spaces we present a method using thermal cameras and Computer Vision (CV) technology, combined with the analytical virtues of Geographical Information Systems (GIS), to track people in urban streets and plazas. The method enables recording of georeferenced positions of individuals in a scene 30 times per second with a spatial accuracy about 25-50 cm. This allows for the analysis of behavior and attendance at a fine scale compared to other established methods for pedestrian behavior monitoring [1]. The use of thermal cameras has the advantage over normal cameras that they can operate independent of light, and in many situations they perform better with Computer Vision software as segmentation of moving objects is easier in thermal video. At the same time concerns for privacy issues when tracking people can be neglected since the identity of individuals cannot be revealed in thermal images. Thus the technique ensures privacy by design. Furthermore the prices on thermal cameras continue to be lowered at the same time as the resolution keeps improving [2]. This add to the practical applicability of such sensors for pedestrian behavioral studies.

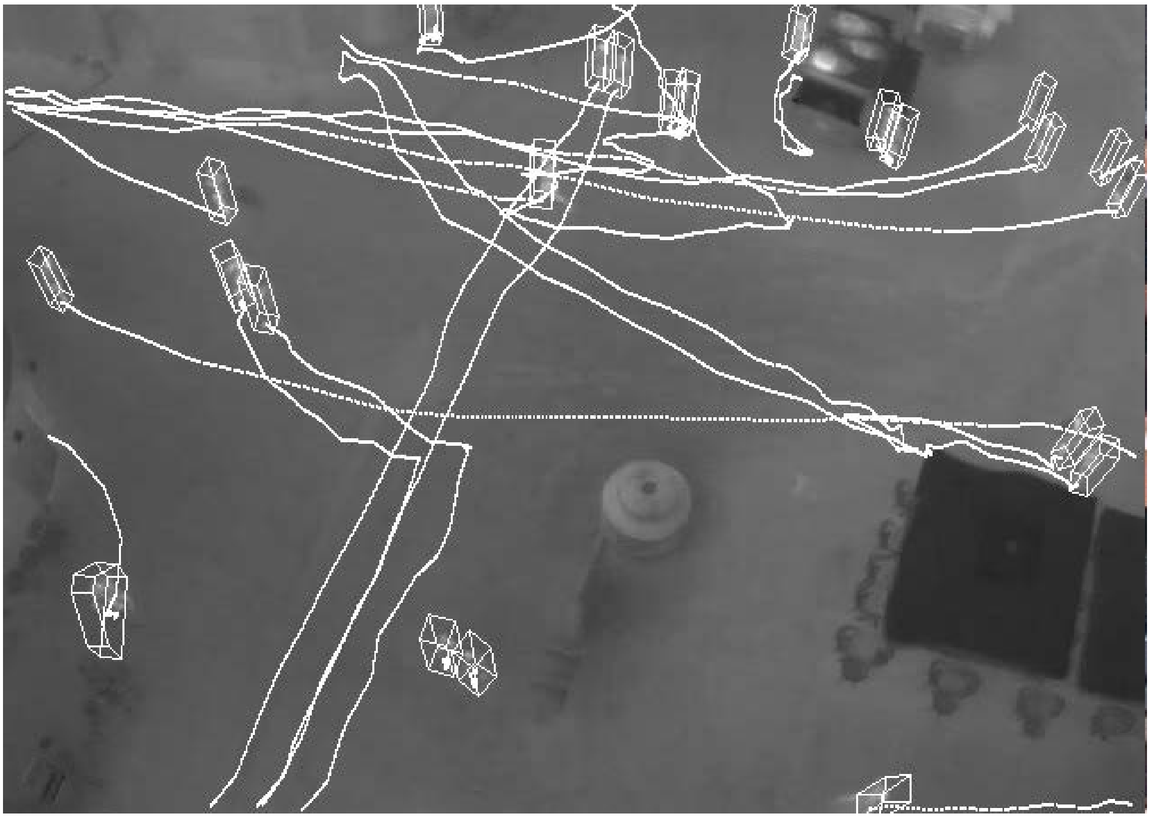


Figure 1. Themal image with tracks of pedestrians from our pilot study.

Our method builds on previous work by [3, 4] and extends the analysis to the GIS domain by capturing georeferenced tracks. This allows for analysis of the tracks in relation to other spatio-temporally referenced data. Environmental variables that might influence movement patterns in urban landscapes such as sunny or shaded areas, wind speed, humidity, rain, can be brought in, as well as a 3D model of the scene, or socio-economic and statistical data for the neighborhood in which the tracking is taking place.

In 2013 we conducted a pilot study in Copenhagen in a pedestrian zone with a continuous flow of pedestrians from several directions that needed to negotiate and avoid each other. A single state-of-the-art uncooled thermal camera with a resolution of 640x480 pixels (Axis Q1922), a lens with a focal length of 10 mm, a viewing angle of 57^o, and 30 fps camera frame rate was used. Background subtraction was applied to detect people. To assess the quality of the trajectories generated by the CV software, a sample of Ground Truth (GT) trajectories were digitized manually for all individuals simultaneously present in the scene in parts of the video recorded. The manual digitization was done in the T-Analyst software developed at Lund University [5].

Tracks of people walking alone or in social groups of different sizes were recorded, as well as people waiting, people having a conversation, and people dragging their bikes or pushing prams or wheelchairs. The tracks of 'facers' working for a charity organization trying to stop people in the street to make them donate to the cause were also recorded in the scene. Our method enables the tracks of individuals in the different situations to be extracted in GIS for further analysis of the detailed movement behaviors in the specific contexts.

Our plans are to carry out a full scale study with video streams from multiple cameras to enable tracking over an urban plaza and adjacent streets for a sustained period of time. The study is to be combined with sampled periods where qualitative data from human observations on the street level will be collected. Further research will be to develop advanced methods in GIS to enable extraction of behavioral parameters for different classes of tracks that can be used to calibrate models of pedestrian movement.

Our approach to tracking urban public life should be seen as a supplement to the traditional qualitative and intuitive manual approaches to collection of data used in studies of urban public spaces and qualities [6,7]. It is the aim that our approach can contribute to the development of new digital methods in this field.

References

[1] Millonig, A,. Brändle, N., Ray, M., Bauer, D., Van der Spek, S. (2009). *Pedestrian Behaviour Monitoring: Methods and Experiences*.

[2] Gade, R., & Moeslund, T. B. (2013). Thermal cameras and applications: A survey. *Machine Vision and Applications*, ***25***, 245–262.

[3] Gade, R., Jørgensen, A., Moeslund, T.B., (2012), Occupancy analysis of sports arenas using thermal imaging, Proceedings of the International Conference on Computer Vision and Applications.

[4] Poulsen, E.S., Andersen, H.J., Jensen, O.B., Gade, R.,Thyrrestrup, T., Moeslund, T.B, (2012), Controlling urban lighting by human motion patterns - results from a full scale experiment. *ACM International Conference on Multimedia (MM).*

[5] Laureshyn, A. (2013). T-Analyst software. http://www.tft.lth.se/video/co_operation/software/, Accessed 1th of July 2014

[6] Gehl, J. (2010). *Cities for People.* Island Press, Washington.

[7] Whyte, W. H. (1980). *The Social Life of Small Urban Spaces*. New York: Project for Public Spaces Inc.

Argos Observation System. Computer Program Enhancing Observation of Classroom Proceedings in Primary School

J. Pisarek, M. Modzelewski

Educational Research Institute, Warsaw, Poland
j.pisarek@ibe.edu.pl, m.modzelewski@ibe.edu.pl

Introduction

Among factors that determine students' achievements, the school environment plays a key role. Interactions between students and teachers, communication in a peer group, educational and didactic methods all constitute what might be called a classroom climate, which is fundamental for school effectiveness analyses [4, 6, 9]. Interactions between students and teachers represent an important mechanism for the development of children's self-regulatory abilities, influencing significantly their involvement in the learning process [12]. The previous research points to three domains of classroom interactions that are crucial for children's academic achievements: classroom management, classroom climate, and number and type of instruction observed in the classroom [2, 8, 10, 11]. Student outcomes has been found to be related to teachers' instructional strategies, i.e., clarity of rules, management of students' behavior, time, and attention [5]. Classroom climate, (i.e., a positive and warm teacher–child relationship characterized by the teacher's regard for children's perspectives and responsiveness to their needs and interests) promotes development of academic skills [1, 3]. The same applied to high quality of instruction (i.e., active monitoring and scaffolding of children's learning and thinking, and tailoring the intruction) [3,7,8,9].

The analysis of the processes occurring in the classroom results in the large amount of information, help to construct programs dedicated for teachers and educators. The observational method is very frequently used in the analysis of the classroom atmosphere.

The need for observational data

Educational Research Institute conducts the Longitudinal School Effectiveness Study. The aim of the study is to identify the key school factors determining the academic achievement of students in Poland. Among many measurements planned during the study, aimed at describing individual, family and school characteristics, the observational study was designed to gather the very needed data on teachers' practices and classroom environment. The research team faced two circumstances that pushed it toward the construction of the computer application for classroom observation, Argos. The first one was the need for reliable description of the typical language (Polish) and maths lessons in large sub-sample of classrooms participating in the study. The second one was the lack of observational tools on Polish market that would not require the presence of a camera during observation. The basis of the theoretical concept used for the observational study and Argos observation system were two acclaimed systems for classroom observation. The systems CLASS (The Classroom Assessment Scoring System) and inCLASS (Individualized Classroom Assessment Scoring System) were created by the team of specialists from the Virginia University, under the supervision of professor Robert C. Pianta.

There were two stages of the observational study, conducted by Educational Research Institute. The first was to collect data that would allow for the examination of the psychometric properties of the Argos observation system. The second was to describe class climate in a selected sub-sample of fifth-grade classrooms participating in the Longitudinal School Effectiveness Study. The observations were conducted from March to June 2013. First, the pilot study was conducted. During this study, 37 observers performed 8 hours of observations in twenty third-grade classrooms, followed by 10 hours of observation in twenty fifth-grade classrooms in the same schools (5 hours during Polish lessons, 5 hours during Math lessons). The main observation study was conducted in 69 fifth-grade classrooms participating in the Longitudinal School Effectiveness Study. 43 observers performed 15 hours of observation in each classroom (8 hours during Polish lessons and 7 hours during Math lessons).

Reliability analysis

For reliability testing of Argos observational system two research designs were used. First one consisted of a number of observational sessions, where all observers participating in a study were asked to code simultaneously the same set of video-recorded lessons, which were coded earlier by the group of experts. The second one employed the double-coded observations (i.e. pair of the observers coded the same lesson in the same mode) in the field during the pilot study. The data were later analyzed with the use of Signal Detection Theory (SDT) and The Receiver Operating Characteristic (ROC) method, which allowed for estimation of sensitivity and specificity of each observer, using dedicated computer application with graphical user interface created in R environment.

Argos as a computer application

Argos is an application developed for Windows operating system. Graphical user interface consists of Main Window and the multiple Tool Windows, situated next to the Main Window. The Tool Windows can be embedded within the Main Window. Most procedures in the application are executed by drag & drop operations, within one Tool Window or between two such windows. The application Argos enables observers to use keyboard shortcuts and the console as well.

The Argos application uses the collections of .xml files, called templates or projects, that specify what is being observed. The modular structure of projects allows easy adaptation of existing templates to suit the researcher needs as well as creating entirely new ones. The following sections describe the projects developed for classroom observation, a part of Longitudinal School Effectiveness Study.

The Argos application works in one of two modes. The first one is a configuration mode, where the observer prepares his working environment. In this mode, observer creates a new project or opens the one prepared earlier and according to the situation of particular class – defines the project elements (e.g. creates students list, defines seating arrangement of students). The classroom is represented in the Classroom Map window as a square grid. Each square can correspond to only one person, represented by circle with a personal identifier inside. Once the observer defines all the necessary project elements, he is allowed to activate the Observation mode.

The Argos observation system

The Argos observation is a very flexible system, which – at this moment – enables classroom observation using three templates: observation focused on a teacher, observation focused on a student and didactic (instructional) observation.

In each of these projects observer registers students' or teachers' activities during lesson by marks events or threads in the rows of Course of the lesson window. Event has a temporary character, e.g. individual behavior or gesture. The application automatically remembers the time for every event. It is possible to define an initiator and the recipient for an event. Threads describe longer states, e.g. lesson stages. Events and threads can have attributes, e.g. teacher work form described for particular lesson stage. In teacher- and student-focused observation the time-sampling methodology was applied, supported by the functionality called the Counter.

During 25 (teacher-focused observation) and 24 (student-focused observation) time samples, the observer registers specific behaviors (from the pre-defined list of observed behaviors) of a teacher or students and records them on the time axis in one of the application windows. A single time sample lasts 1.5 minutes – during 30 seconds the observer is observing a teacher or a student, and during the subsequent 60 seconds he or she marks the observed behavior on the time axis. The Counter automates the whole sampling process, enabling the observers to focus only on students' or teacher's behavior.

The main Counter's task is to countdown the time within the time samples and separating the observation and registration stages. After the time sample dedicated for observing one students ends, the Counter automatically highlights the next one on the Classroom map, helping the observer to follow the student sampling design. Additionally, in order to separate observations of particular students, the Counter automatically switches the active row in the Course of the lesson window. The data from the observation are saved as an xml file.

References

[1] Birch, S.H., Ladd, G.W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology*, ***35***, 61-79.

[2] Connor, C.M., Morrison, F.J., Katch, E.L. (2004). Beyond the reading wars: The effect of classroom instruction by child interactions on early reading. *Scientific Studies of Reading*, ***8***, 305-336.

[3] Connor, C.M., Son, S.-H., Hindman, A.H., Morrison, F.J. (2005). Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*, ***43***, 343-375.

[4] Early, D.M., Maxwell, K.M., Burchinal, M., Alva, S., Bender, R.H., Bryant, D., Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, ***78***, 558-580.

[5] Emmer, E.T., Stough, L. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, ***36***, 103–112.

[6] Hamre, B.K., Pianta, R.C., Mashburn, A.J., Downer, J.T. (2007). Building a science of classroom application of the CLASS framework in over 4000 U.S. early childhood and elementary classrooms. Available online: http://fcd-us.org/sites/default/files/BuildingAScienceOfClassroomsPiantaHamre.pdf [Last accessed on November 17, 2014]

[7] Juel, C., Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*, ***35***, 498–492.

[8] La Paro, K.M., Pianta, R.C., Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, ***104***, 409–426.

[9] Mashburn, A., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O.A., Bryant, D., Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, ***79***, 732–749.

[10] Pianta, R.C., Mashburn, A.J., Downer, J.T., Hamre, B.K., Justice, L.M. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, ***23***, 431–451.

[11] Stipek, D.J., Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, ***19***, 375-397.

[12] Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA, USA: Harvard University Press.

Measuring Equilibrium and Motor Learning Deficits during Aging with the Elevated Unsteady Board in Mice with Precocious Cerebellar Degeneration

Pascal Hilber

Psychology and neuroscience of cognition and affectivity laboratory
Psy, nca EA 4700
UFR Sciences, University of Rouen, Mont saint Aignan, France

Abstract

Lurcher mutant mice are well known to exhibit massive and neonatal degeneration of Purkinje cells in the cerebellar cortex which influences normal ageing process. In spite of their severe motor disturbances, at 3 months old, these mice are still able to learn a motor task. We evaluated such abilities during aging with the Elevated Unsteady Board test. The results clearly showed that the precocious and focused degeneration of Purkinje cells have mild and long term consequences on motor learning functions in these mutants.

Introduction

Many tests exist to investigate motor abilities in rodents. Some of them aim at measuring motor coordination and equilibrium abilities when the animals are free to move (i.e. the hole board and the parallel rod floor tests). On the contrary on other tests animals are forced to adjust their posture in response to a moving surface (i.e. the rotorod test). Nevertheless, in these entire tests animals can move to maintain balance and equilibrium evaluation is made on dynamic conditions. The Elevated Unsteady Board was created in our lab to measure equilibrium maintenance in response to a moving surface but on static condition (Hilber, Lalonde, & Caston, 1999): only movement of the animal could provoke equilibrium disturbance and only adapted and quick postural body adjustment permit to avoid the equilibrium loss and fall. This test permitted to show that the cerebellar Lurcher mutant mice exhibited equilibrium disturbances on static conditions. These mutants are well known to exhibit complete degeneration of Purkinje cells in the cerebellar cortex (Vogel, Caston, Yuzaki, & Mariani, 2007) which begins very early in development stages and influences normal ageing process (Hilber & Caston, 2001). Here we used the Elevated Unsteady Board apparatus to investigate the effect of daily training on equilibrium abilities in the cerebellar Lurcher mutant mice during aging.

Material and protocol

Animals

The mice were obtained in our laboratory and reared in standard conditions (12h light (8am-8pm)- 12h dark (8pm-8am), 21°C, food and water available *ad lib*. Heterozygous Lurcher (+/Lc) were obtained crossing females of wild type of B6CBA strain (+/+) with heterozygous Lurcher males of the same strain. 80 animals were tested, split in four groups of ages (3, 9, 15 and 21 months), containing 10 +/+ and 10 +/Lc.

Description of the elevated unsteady board

As shown in the photo near to the Figure 1, the unstable board was circular (diameter, 8.5 cm), made of aluminium, and set at a height of 1 m from the floor. A foam rub mat was disposed on the floor to cushion animal's falls. The surface of the platform was covered with a plastic sheet for the purpose of preventing excessive sliding. The platform was positioned on a ballpoint, causing a tilt of 30° in any direction whenever the mice strayed from the middle part. The minimal weight necessary to incline the platform downward was 4g.

Protocol

At the beginning of the test, each mouse was placed in the middle of the platform, which was in a horizontal position. Thus, only motions of the animal could provoke tilting of the platform and only adapted repartition of muscular strength in the limbs and the body could permit the mouse to restore equilibrium and to maintain balance. The trial ended when two or more paws of the mouse were out of the circumference of the platform or when the mouse reached the cut off period (180 sec.). Each mouse was subjected to 3 trials per day. The intertrial interval was 15 min. The mice were trained daily until they reached the learning criterion fixed at three consecutive trials of 3 min. If they didn't reach this criterion after 10 days, the training was stopped. During each trial, we measured the time until falling. For each animal, the scores of the 3 daily trials were averaged to get the mean score per day. Then, we calculated the mean daily score (+/- sem) obtained by each group of mice. The mean scores were given +/- SEM ($\sigma/\sqrt{n}$).

Results

Three way ANOVA and appropriate post hoc comparisons with Tukey HSD (High Significant Difference) test were used for the purpose of statistical analysis. In all instances, $p<.05$ was considered as significant. 2x4x10 ANOVA (2 genotypes, 4 groups of age, 10 days, with repeated measures on the last factor) revealed a significant training [$F(9,648) = 34.2$; $p<0.001$], genotype [$F(1,72) = 565.23$; $p<0.001$] and an age effects [$F(3, 72) = 10.1$; $p<0.001$] with significant age x genotype x repeated measures interaction [$F(27, 648) = 7.3$; $p<0.001$], which indicates that the evolution of the performances during training differed for the two genotypes as a function of age (see Figure 1). Whatever the age, the scores performed by the mutants during the first day that is before training were lower than those of controls ($p<0.001$). Whereas the scores of +/+ mice significantly decreased between 3 and 15 months ($p<0.001$), no significant decrease of performances was observed in Lurcher mice with age: concerning the scores performed during the last day (deficits after training), the scores of +/Lc mice were always lower than those of controls whatever the age ($p<0.05$ at 3 months and $p<0.001$ at 9, 15 and 21 months). The performances did not decrease with age in +/+ mice (they were maximal at all ages) but decreased significantly in the +/Lc mice as soon as 9 months (3 versus 9 months: $p<0.001$).

Discussion

The Lurcher mutants are an interesting tool to observe the consequences of a focused neurodegeneration on ageing processes (Hilber & Caston, 2001). The results of this study showed that in spite of their equilibrium disturbances on static conditions the young mutants were still able to elaborate and execute a novel equilibrium motor task on static conditions. Nevertheless although the latency before falling from the platform improved with training, the scores of the mutants were always poorer to those of controls, indicating motor learning deficits in the mutants. Equilibrium abilities evaluated before training decreased in +/+ mice indicating age-related impairments. Such impairments were not observed in +/Lc mice certainly because of a floor effect. In spite of their equilibrium deficits, all the +/+ mice reached the learning criterion, indicating that if their motor abilities were affected by the age, their motor learning abilities were not: they were still able to elaborate and to execute a motor program which permitted them to stay immobile on the unstable platform to avoid the fall. The results are on the opposite in the Lurcher mice since, as mentioned above, their equilibrium abilities were not significantly deteriorated but their motor learning decreased as soon as 9 months old. Thus, with age, these mice became progressively unable to elaborate and execute a novel motor program adapted to the task: on the rotorod the walking strategy disappeared (Hilber & Caston, 2001) and, on the unstable platform, they became unable to use muscular synergies which would permit to stay immobile on the apparatus.

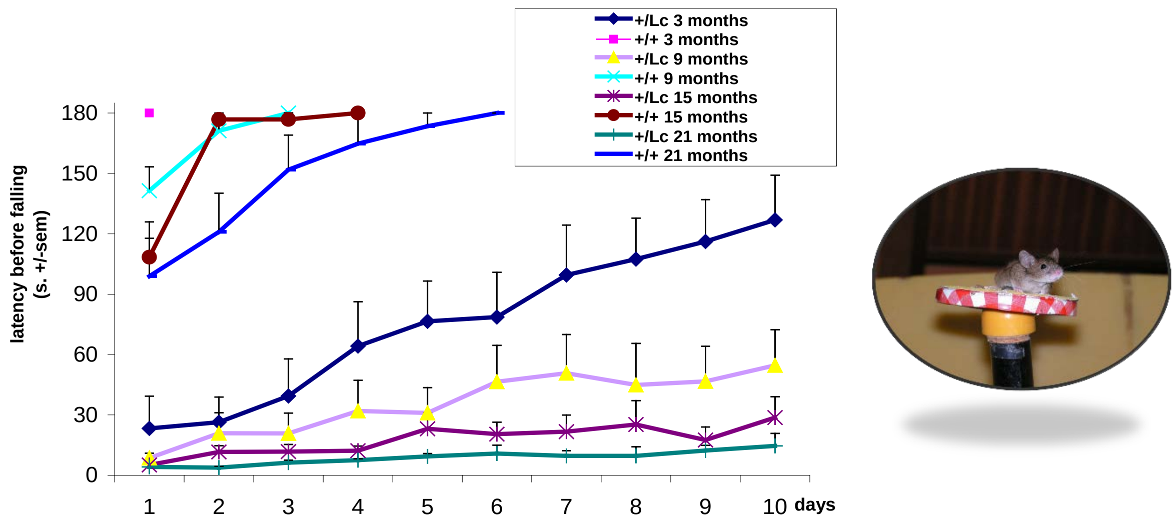


Figure 1. Evolution of the performances (latency before falling) of +/+ and +/Lc mice on the Elevated Unsteady Board during training (days).

Ethical statement

All the experiments reported here were conducted in accordance with the ethical recommendations of the Directive 2010/63/EU of the European Parliament and of the Council for the protection of animals used for scientific purposes.

References

[1] Hilber, P., & Caston, J. (2001). Motor skills and motor learning in Lurcher mutant mice during aging. *Neuroscience*, ***102****(3)*, 615–23.

[2] Hilber, P., Lalonde, R., & Caston, J. (1999). An unsteady platform test for measuring static equilibrium in mice. *Journal of Neuroscience Methods*, ***88****(2)*, 201–5.

[3] Vogel, M. W., Caston, J., Yuzaki, M., & Mariani, J. (2007). The Lurcher mouse: fresh insights from an old mutant. *Brain Research*, ***1140***, 4–18.

Effects of Clozapine, Aripiprazole and Bitopertin in Rats Social Withdrawal Assessed in an Automatic Social Interaction Test

S. Deiana, H. Rosenbrock and R. Arban

Department of CNS Diseases Research, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach, Germany. serena.deiana@boehringer-ingelheim.com

Introduction and aims

Deficient N-methyl-D-aspartate (NMDA) signaling is hypothesized to underlie cognitive and negative symptoms associated with schizophrenia which to date are poorly responsive to currently available antipsychotics.

In healthy volunteers, the NMDA receptor antagonists phencyclidine (PCP) and ketamine induce psychotomimetic effects mirroring positive, negative and cognitive symptoms of schizophrenia [1, 2]. Moreover, these NMDAR antagonists can exacerbate positive and negative symptoms of schizophrenics [3-5]. Social withdrawal is one of the negative symptoms of schizophrenia and it can be mimicked in animals via PCP administration [6,7].

In the present set of experiments, we aimed at implementing an automatic three-chamber social interaction test in rats. Social withdrawal was induced by PCP treatment [7] and possible reversal by the antipsychotics clozapine, aripiprazole and the GlyT1-inhibitor Bitopertin [8, 9] was evaluated. The development of a robust pre-clinical model that closely mimics social interaction deficits typical of schizophrenia would contribute to the discovery of efficacious therapies aimed at alleviating such symptom.

Methods

Male Wistar rats (Janvier Labs; 8-9 weeks old, n=6-13) received either PCP (0.5-3 mg/kg, sc) or saline for three consecutive days; on the third day, rats were injected with PCP, and either vehicle or Clozapine (2.5 - 10mg/kg, po, 1hr prior to test), Aripiprazole (1, 3 mg/kg, ip, 30 min. prior to test) or Bitopertin (0.3, 1, 3 mg/kg, po, 1hr prior to test). On the same day, rats were tested in the social interaction test, similarly to the three-chamber model described by Moy and co-workers [10] in mice. Rats were habituated for 10 minutes to a rectangular arena consisting of three chambers divided by two transparent walls provided with a 10 cm squared passing door. In the left and the right chambers an empty, transparent, holed cubiculum was positioned. Immediately after the habituation session, a juvenile co-gender Wistar rat (3-4 weeks old) was positioned in one of the empty cubicles in a fully counterbalanced fashion. The transparent cubicles permitted visual, olfactory, auditory, and some tactile contact between the stranger and the test rat. The animal’s behavior was recorded with a video-camera connected to a PC, such that a video tracking software (Anymaze, Stoelting) extrapolated x,y coordinates to compute the animal’s position. Locomotor activity and time spent in the 6 cm vicinity area from each of the two cubicles were measured. This 6 cm area was defined via software to determine direct social contacts based on the optimal distance for the tested rat to sniff at a stranger. During the first validation experiment, social interaction time was also manually scored for comparison with software-produced data.

Results

Manually-scored social interaction time was comparable to the times computed by the tracking software. Control rats showed a strong preference for the inhabited cubiculum versus the empty one. PCP induced a robust and dose-dependent reduction of social interaction duration, with doses superior to 2mg/kg inducing motor incoordination and increased activity. Clozapine did not reverse PCP (1mg/kg)-induced social withdrawal, with the high dose inducing sedation. Bitopertin revealed inefficacious at reversing PCP-effects. By contrast, Aripiprazole induced a dose-dependent amelioration of the social withdrawal induced by PCP.

Conclusions

These findings proved that social withdrawal induced by NMDA signaling disruption can be modeled in rats using the three chambers procedure. Pharmacological results are in line with previous findings [7] in rats tested in a dyadic open field social interaction set up, showing that the two models (dyadic versus unilateral) of social interaction test are pharmacologically equally sensitive. Finally, in line with some small studies conducted in humans [11, 12] the social withdrawal was ameliorated by antipsychotic treatment, suggesting that this model may serve as a tool for testing novel drug candidates for the treatment of negative symptoms of schizophrenia.

All experimental procedures were approved by the Ethics Committee and the Regierungspräsidium Tübingen and adhered to the guidelines of the committee for Research and Ethical Issues OASP 1983.

References

[1] Krystal JH, Karper LP, Seibyl JP, et al. (1994). Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans: psychotomimetic,perceptual, cognitive, and neuroendocrine responses. *Archives of General Psychiatry, **51**(3),* 199-214.

[2] Adler CM, Goldberg TE, Malhotra AK, Pickar D, Breier A. (1998). Effects of ketamine on thought disorder, working memory, and semantic memory in healthy volunteers. *Biological Psychiatry, **43**(11),* 811-816.

[3] Malhotra AK, Pinals DA, Adler CM, et al. (1997). Ketamine-induced exacerbation of psychotic symptoms and cognitive impairment in neuroleptic-free schizophrenics. *Neuropsychopharmacology, **17**(3),* 141-150.

[4] Lahti AC, Koffel B, LaPorte D, Tamminga CA. (1995). Subanesthetic doses of ketamine stimulate psychosis in schizophrenia. *Neuropsychopharmacology, **13**(1),* 9-19.

[5] Curran C, Byrappa N, McBride A. (2004) Stimulant psychosis: systematic review. *Britisch Journal of Psychiatry, **185***, 196-204.

[6] Gururajan A, Taylor DA, Malone DT. (2012). Cannabidiol and clozapine reverse MK-801-induced deficits in social interaction and hyperactivity in Sprague-Dawley rats. *Journal of Psychopharmacology, **26**(10),* 1317-32.

[7] Sams-Dodd F. (1995). Automation of the social interaction test by a video-tracking system: behavioural effects of repeated phencyclidine treatment. *Journal of Neuroscience Methods, **59**(2),* 157-67.

[8] Goff DC. (2014). Bitopertin: The Good News and Bad News. *JAMA Psychiatry*.

[9] Bugarski-Kirola D, Arango C, Fleischhacker WW, Bressan R, Nasrallah H, Lawrie S, Blaettler T, Garibaldi G, Reid C, Marder S. (2014). Efficacy and safety of adjunctive bitopertin versus placebo in subjects with persistent predominant negative symptoms of Schizophrenia treated with antipsychotics – update from the Searchlyte programme. *Schizophrenia Research, **153**(1),* 29.

[10] Moy SS, Nadler JJ, Perez A, Barbaro RP, Johns JM, Magnuson TR, Piven J, Crawley JN. (2004). Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes, Brain, and Behavoir, **3**(5),* 287-302.

[11] .Kane JM1, Carson WH, Saha AR, McQuade RD, Ingenito GG, Zimbroff DL, Ali MW. (2002). Efficacy and safety of aripiprazole and haloperidol versus placebo in patients with schizophrenia and schizoaffective disorder. *Journal of Clinical Psychiatry, **63**(9),* 763-771

[12] Potkin SG1, Saha AR, Kujawa MJ, Carson WH, Ali M, Stock E, Stringfellow J, Ingenito G, Marder SR. (2003). Aripiprazole, an antipsychotic with a novel mechanism of action, and risperidone vs placebo in patients with schizophrenia and schizoaffective disorder. *Archives of General Psychiatry, **60**(7),* 681-90.

Pre-processing of electromyography startle data: A novel semi-automatic method

L.R. van Bedaf[1,2], L. Heesink[1,2] and E. Geuze[1,2]

[1]Research Center Military Mental Health Care, Ministry of Defense, Utrecht, The Netherlands
[2]Department of Psychiatry, University Medical Center, Utrecht, The Netherlands

The startle blink response is an eye blink in response to an (acoustic) startle stimulus that is frequently used in psychophysiology research and various clinical settings [1]. There are various methods used to preprocess the data, but these methods are time-consuming, and there is little consensus on which method is best [2]. Here we propose improvements of pre-processing this data.

Human facial surface EMG of the startle reflex is a very dynamic and sensitive measurement. Traditionally, each trial is visually inspected to check whether the measurement is reliable [2]. This method is a very subjective way of checking data and very time consuming. Moreover, when examining a large volume of data, errors are easily made. Visual inspection by two researchers would make it more reliable, however, this is even more time-consuming and the classification would still be subject to subjectivity. Therefore we propose a more objective semi-automatic way of checking the reliability of the trials using the variation of the raw EMG. After rejecting unreliable measurements there is still variation in the baseline across participants, sessions and even within a session due to differences in muscle tension and noise. Therefore, a baseline correction might be important for a good peak amplitude estimation. However, there is no consensus about baseline measurement in EMG experiments [2]. Therefore, we additionally propose an automatic method for determining a baseline.

Data was collected from 37 male combat veterans. The medical ethical board of the University Medical Center Utrecht approved this study. Startle responses were elicited by a sudden acoustic white noise sound of 105 dB while a fixation cross was presented on the screen and were measured by the use of electromyography (EMG). Two electrodes were placed on the skin surface above the orbicularis oculi muscle [1], and the signal was recorded using the Biopac MP150 system with a sampling rate of 1.000 Hz. Data analysis of the raw EMG signal was performed using the software AcqKnowledge version 4.3. The raw EMG signal was filtered offline with two filters. A Finite Impulse Response (FIR) band-pass filter of 28 to 500 Hz (1001 coefficients tapered with Hamming window) was used to filter out the high and low artifacts (such as motion artifacts). This frequency range has a maximum trade-off between suppressing artifacts and retaining the true EMG signal [2, 4]. Next an Infinite Impulse Response (IIR) band stop filter (Frequency 50.38 Hz, Q 30.59) was used to filter out electrical noise [3]. A frequency of 50.38 Hz was used instead of the standard 50 Hz because this was the dominant frequency in our study and a filter centered at 50.38hz was more effective. We used a Q of 30.59 to keep the discarded bandwidth as low as possible. From this signal the root mean square (RMS) was calculated with a sliding mean of 0.3 seconds [5].

For a reliable measurement, an interval of 150 ms before the time window of the startle responses should be stable. Variation caused by, for example, spontaneous blinking, eye movement and disturbance in the lines can contaminate the peak measurement. To check the stability, the variation based on the standard deviation (SD) of the raw filtered EMG signal was calculated in this interval. If the variation of an interval rises above the threshold (0.0115 mV), it is regarded as unstable and therefore the trial was rejected. If the variation stays below the lower threshold (0.0085 mV) it can be assumed that the variation is negligible and the measurement is reliable. Surface EMG is a very sensitive and dynamic measurement and every threshold is to some extent arbitrary, therefore, we propose a semi-automatic method in which the trials between the lower and the upper threshold will be manually classified by visual inspection by two researchers.

We performed both the traditional and the new semi-automatic method on 200 trials. All trials were classified automatically based on the SD and by 2 examiners that compared and discussed their decision till consensus was reached. The classification by 2 examiners was used as the “golden standard” of reliable and unreliable trials. The semi-automatic classification performed well in 200 trials; 99.5% of the trials were correctly classified and only 5% of these trials were classified for further manual inspection.

Other semi-automatic methods for pre-processing of EMG signal are known [6], but these methods focus on detection of the on- and offset of an event rather than checking stability of the baseline. Furthermore, none of these methods are applied to startle eye blink data. Although some of these methods (for example based on wavelet transformation [7]) could, in theory, do the same as the method proposed here, but are much more complex and as the current method proves, this complexity is not necessary for a good classification.

In addition, we propose an automatic method for baseline calculation. The longer the interval of the baseline calculation, the closer the value will be to the true mean. However, this increases the risk of including artifacts such as blinking or eye rolling, thus leading to a higher baseline estimation and therefore a lower peak value. To correct for this effect we calculated the mean of the baseline activity within the baseline interval of 1 second in epochs of 100 ms (10 epochs per baseline, calculated on the RMS data before the startle response window). If in the baseline interval the variation of an epoch was higher than the upper threshold (0.0115mv), this epoch was excluded in the baseline calculation. This baseline correction gives us the advantage of a large interval while constraining the risk of including artifacts. These baseline values were subtracted from the peak values.

In conclusion, we propose a semi-automatic objective method of checking the startle measurements. This can be performed automatically and so decreases subjectivity in the classification, human error and time. Furthermore, an automatic baseline correction method is proposed by using a long interval in which artifacts are removed. With this improved semi-automatic method, we try to increase the quality of startle EMG data.

References

[1] Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology, **42**(1)*, 1-15.

[2] Van Boxtel, A. (2010). Filters for optimal smoothing of acoustic and electric blink reflex EMG responses to determine blink response magnitude. *Biological Psychology, **85**(2)*, 299-305.

[3] Zschorlich, V.R. (1989). Digital filtering of EMG-signals. Electromyogr. *Clinical Neurophysiology, **29**(8)*, 81-86

[4] Van Boxtel, A., Boelhouwer, A.J., Bos, A.R. (1998). Optimal EMG signal bandwidth and interelectrode distance for the recording of acoustic, electrocutaneous, and photic blink reflexes. *Psychophysiology, **35**(6)*, 690-697.

[5] De Luca, C.J. (1997). The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics, **13**(2)*, 135-163.

[6] Merlo, A., Farina, D., Merletti, R.A. (2003). Fast and reliable technique for muscle activity detection from surface EMG signals. *IEEE Transactions on Biomedical Engineering, **50**(3)*, 316-323.

[7] Reaz, M., Hussain, M., Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: Detection, processing, classification and applications. *Biological Procedures Online, **8**(1)*, 11-35.

Cued fear conditioning:

Minimizing ‘contextual leftover’-freezing by Maximizing changes of context

H. Van Craenendonck[1], L.Ver Donck

Neuroscience Drug Discovery, Janssen Research and Development, Janssen Pharmaceutica NV, Beerse, Belgium

[1]hvcraene@its.jnj.com

Introduction

Rodent fear conditioning is a relatively simple, cognitive-based paradigm extensively used to study the neurobiological mechanisms underpinning associative and emotional learning and memory. In the cued fear conditioning, animals learn to predict aversive events by associating a neutral stimulus (Conditional stimulus, CS: white noise) with an aversive stimulus (Unconditional stimulus, US: electrical foot shock). This results in the expression of fear responses when later re-exposed to the original neutral stimulus in a different context (recall). The dependent measure used as a read-out for memory function is a characteristic freezing response to the CS in rodents. This response is defined as the absence of any movement, except that required for breathing.

Ideally, rodents exposed to such novel context would not display a freezing response in the absence of the noise (CS). However, in our original setup, a certain amount of freezing was still observed prior to exposure to the noise during the recall session, which indicated that the context differences were not optimal. This ‘contextual leftover’-freezing possibly could compromise drug effects because the window between baseline and CS-response becomes smaller.

The objective of this study was to eliminate factors contributing to this ‘contextual leftover’-freezing response during recall.

Methods

Subjects

Male C57Bl/6 mice (22-35 gram at test) were obtained from Janvier (France). Mice were single housed on arrival and allowed 5 days of acclimatization before starting the experiment. The Institutional Ethical Committee on Animal Experimentation approved the experimental protocols, in compliance with Belgian law (Royal Decree on the protection of laboratory animals dd. April 6, 2010) and the facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).

Apparatus

Fear conditioning studies were performed in mouse modular test chambers (MED Associates, USA) placed in ventilated, illuminated cabinets. The modular test chamber was equipped with a speaker controlled by a white noise amplifier. Each chamber was equipped with a top-view video camera.

Conditioning. A white PVC test cubicle (13×13×13cm, Context A) was placed in the center of the test chambers. The foot shock was delivered through stainless steel floor grids which were calibrated for each cage to ensure appropriate shock intensity. All programming, timing, and shock presentations were computer controlled (Med-PC IV software). Freezing responses were determined using an automated video acquisition based system (Activity detection module, EthoVision XT9®, Noldus, The Netherlands).

Recall. For the auditory CS test, contextual cues were changed. In the standard setup (Context B) the aluminum wall panels, door and back panel, were replaced by black and white striped PVC panels. The grid floor was covered with an opaque grey plastic plate and 1 ml of an aromatic substance (vanilla) was added to the waste pan. It appeared that these changes to the context were insufficient (figure 1): the test box cover was identical, the aluminum columns between the wall plates were still visible, the white PVC-plate on the bottom didn't hide the grid floor completely, and the box shape and illumination remained the same.

Criteria for the new context development therefore where: different shape and color, different lighting, absence of odor usage and optimal environmental conditions for automated video analysis.

In the modified context C, the grid floor was removed and a new insert (grey PVC cylinder, Ø 12 cm, H=13 cm) was placed on top of a light box made of black infra-red transparent PVC and equipped with near infra-red LED's. Four visual light LED's were integrated in the cylinder. The insert was covered by an infra-red transparent lid.

Procedure

On the conditioning day, mice were placed in the test chamber (context A) and after 2 minutes of free exploration, a 75 dB white noise serving as the conditioned stimulus (CS) was presented for 30 sec. Then a foot shock (0.70 mA, 1 sec) was applied, which served as the unconditioned stimulus (US). This presentation of CS/US pairings was repeated at, 240 and 360 seconds after the start of the conditioning phase. The mouse was removed 90 seconds after the last pairing and returned to its home cage.

Twenty-four hours later, each mouse was returned to the test chamber in which the environmental and contextual cues were changed to context B or C. The mice were placed in the box and freezing behaviour was determined for 2 minutes in the absence of the auditory CS.

Results

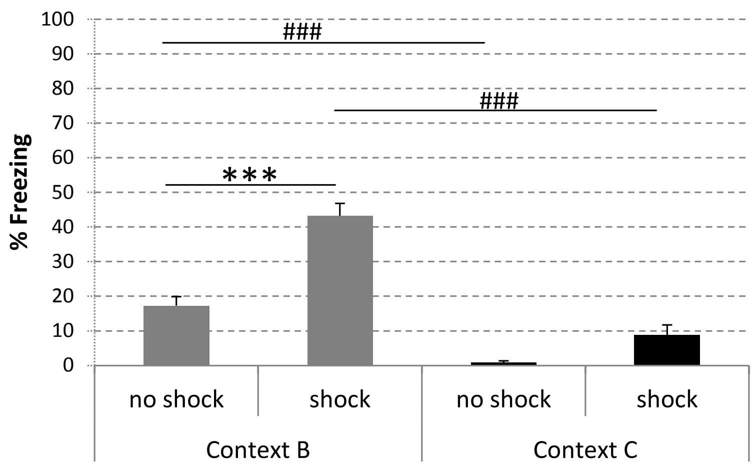


Figure 1. Percent of time freezing ± SEM during the first 2 min (pre-cue) in a novel context at recall. Context B: n=8 and Context C: n=12 per group (see text for details).

One-way ANOVA with context as a factor showed a significant difference between contexts ($P< 0.0001$).

The effect of context was shock-dependent (2-way ANOVA, context x shock: $P=0.0012$). Shocked mice in the modified context C showed lower % freezing than those in the standard context B ($P<0.0001$ – Bonferonni-adjusted).

Discussion

In our standard setup (context B), we observed a substantial amount of freezing during the baseline period prior to exposure to the tone during recall. We hypothesized that this ‘contextual leftover’-freezing could possibly compromise compound effects because of a smaller window between freezing elicited by the CS vs. baseline.

The effort to enlarge the difference between the conditioning context and the test context resulted in a much smaller ‘contextual leftover’-freezing and a larger difference compared to the CS induced freezing. It also resulted in an improved environment for automated video analysis using EthoVision XT9®, in terms of image contrast, reflection, illumination, etc...

In our opinion, it will be very hard to further decrease the ‘contextual leftover’-freezing, because the influence of other factors, such as the experimenter, handling, lab environmental conditions, is very difficult or even impossible to exclude.

Validation of automated detection of impaired motor function of rats on the rotarod: Use of EthoVision XT®.

Michel Mahieu*, Roland Willems and Luc Ver Donck

Janssen Research and Development, Bierse, Belgium
* mmahieu@its.jnj.com

Introduction

The rotarod is used to evaluate whether test-compounds affect motor performance of a rodent, or induce fatigue. Motor impairment in humans can be a disturbing side effect with a large impact on the patient's quality of life as the patient may no longer be authorized to drive or work with machines. The principle of this test is that rats or mice are first trained to walk on a rod rotating at a certain speed. Once the animals have learned this, the effect of a test-compound on their motor performance is evaluated. Animals experiencing impaired motor coordination are unable to cope with the rotating rod and will drop off when the rotation speed exceeds their motor coordination capacity. The more impaired the animals are, the sooner they fall off the rod. However, mice show a particular coping behaviour on the rotarod when experiencing motor coordination problems: to prevent dropping off, mice can grip themselves to the rod and turn around without falling off. This results in late (or no) falling off the rotarod, which would incorrectly indicate that the compound tested did not disturb motor function. The number of times mice turn around on the rotarod therefore represents a secondary measure of disturbed motor coordination. Another secondary parameter to measure impairment, in mice is the time animals spent at the back of the rod [1].

Methods

The Institutional Ethical Committee on Animal Experimentation approved the experimental protocol, in compliance with Belgian law (Royal Decree on the protection of laboratory animal's dd. April 6, 2010) and the facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). Male Spraque Dawley rats (body weight 240-300 g, Harlan Italy) were habituated to the environment for 1 week. Rats were first trained to walk on the rotarod (Med Associates) in 5 min long sessions every 30 min at a constant rotation speed of 8, 12 and 16 rotations per minute (rpm) respectively. Rats were placed back on the rod each time they fell off, until the 5 min session was completed. Next a session was performed with the rotarod rotating at an incremental speed starting at 3 rpm and gradually accelerating over a 5 min period up to 30 rpm. The next day, a dose of the test-compound was administered, and rats were tested at 30 min intervals in 5 min sessions at accelerating speed. A high resolution video camera (Samsung SDC-435) was positioned 45 cm above each lane of the rotarod to record the sessions. EthoVision XT9® (Noldus) was used to analyse the video images. The background was covered with black plastic material such that a white rat could be readily determined as an object in EthoVision®. A region of interest area was identified around the rod and this area was divided in 3 separate zones used for detection of turnaround behaviour (Figure 1). The mid zone was positioned on the rod with front and back zone alongside. A fourth zone named reference zone (RZ) was defined next to the back zone. The centre point of gravity (COG) of the object (rat) was determined, as well as its occurrence in one of the zones and its distance to the RZ. The total time the rat stayed on the rod (latency time) was determined by the time until the COG was not detected anymore. The sequential transition of the COG between different zones was determined to indicate the turnaround manoeuvres of the rat using a custom made macro in MS Excel: a sequential occurrence of COG in the mid zone followed by the back zone, the front zone and the mid zone again was defined as a turnaround event of the rat. The measures of the distance of the COG to the RZ were used to calculate the time spent by the rat in the back of the rod: a distance <6.8 cm was considered as the rat spending time in the back. This cut-off was based on the rear quarter part of the mid zone (i.e. 1.8 cm) + 5 cm length of the back zone, marking the total distance to the RF. The total time of the COG spent <6.8 cm distance from the RZ was calculated and expressed relative the total latency time of the rat on the rod.

Haloperidol, diazepam or vehicle was administered by SC route, immediately after the first session on the test day.

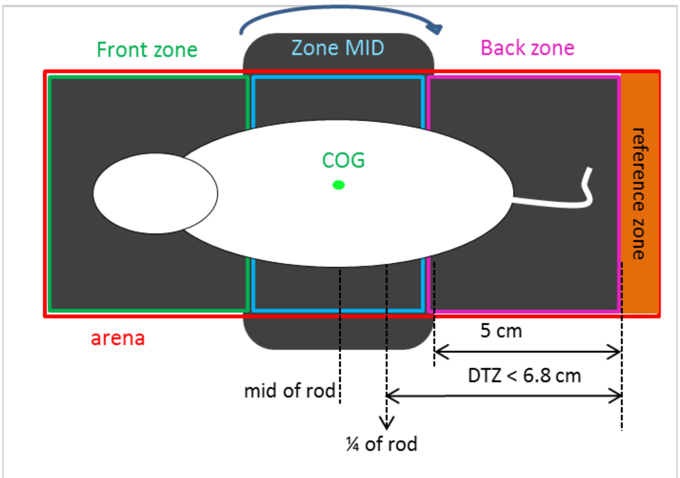


Figure 1. Top view image of one lane of the rotarod. Overlays show the area of interest (red) and the 3 individual zones (green, blue and pink) covering the rod of the rotarod (MID), the front and back zones as determined in EthoVison®. Zones are used to calculate the number of turnarounds. The reference zone is used to calculate the time in the back and is defined at 5 cm from the edge of the rod. The arrow indicates the turning direction of the rod.

Objective

The objective of the present study was to validate these secondary endpoints for performance of rats on the rotarod.

Results

Vehicle treated rats performed well on the rotarod with latency times ranging between 245±19 and 300±0 sec (n=10/group) over the course of the test phase. The time spent in the back ranged between 1.0±0.1 and 3.3±2.3% of total latency time. Rats did hardly make any turnarounds. This is most likely due to their much higher body weight compared to mice which makes it very difficult for them to remain clamped to the rotating rod. This parameter will therefore not be used further in this study. Both drugs induced a dose-dependent reduction in latency time: haloperidol from 0.4 mg/kg onwards and diazepam from 1.25 mg/kg onwards. Time spent in the back increased in parallel with decreasing latency times for both drugs. Rats treated with haloperidol spent up to 18.4±2.3 % of time in the back at the highest dose of 0.31 mg/kg tested. Diazepam (10 mg/kg) treated rats displayed up to 8.0±1.5 % of time in the back.

Conclusion

With advanced automated measurements using image analysis software we are able to define new parameters that relate to the performance of rats on the rotarod. However, with the drugs tested so far we did not find a differentiation between the time spent in the back and the latency time. The turnaround parameter does not appear to be relevant for rats because they are too heavy to turn around. This is in contrast with mice, which use this strategy to cling to the rod so that there was a difference between latency and turnarounds for drugs tested [1]. Testing of other pharmacological classes is needed to evaluate the added value of these additional parameters in rats.

Reference

[1] Mahieu M., Willems R., Hoekstra L., Ver Donck L. (2012). Automated detection of aberrant behaviour of mice on the rotarod: Use of EthoVision XT®. In *Proceedings of Measuring Behaviour 2012: 8th International Conference on Methods and Techniques in Behavioural Research*, 28-31 August, Utrecht, The Netherlands. pp. 431-433. Wageningen, The Netherlands: Noldus IT.

Characterization of Burrowing Deficits in the MIA Model of Osteoarthritis

L.A. Bryden, N. Gorodetskaya, J.R. Nicholson and A. Pekcec

Department of CNS Diseases Research, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach an der Riss, Germany

Burrowing deficits as a measure of pain-related behaviour

Currently the assessment of pain-related behaviours in preclinical animal models relies on readouts that assess pain evoked by an external stimulus. These assays measure spinally-mediated reflexes that may not encompass supraspinal modulation of pain signalling and a number of reviews have highlighted the apparent lack of translation of these evoked measurements to the clinical setting [1,2]. Burrowing is an innate and self-rewarding behaviour naturally displayed and pain-induced deficits are a sensitive indicator of a reduction in animal well-being. These deficits are a specific read-out for pain in preclinical inflammatory and neuropathic pain models and are sensitive to reversal by analgesics, which validate the utility of this behaviour as an addition to current assays [3,4]. From a practical perspective, measuring burrowing behaviour does not require the animal to be restrained, as in some currently used evoked assays, which reduces the impact of stress as a confounding variable. Additionally, the lack of experimenter input during testing means burrowing is an objective assay that eliminates any possible experimenter bias. The injection of monosodium iodoacetate (MIA) into the rat knee joint induces osteoarthritis (OA)-like lesions with associated pain-related behaviours [5]. The aim of our study was to evaluate burrowing deficits in this preclinical model of OA and to establish if these deficits were reversible by analgesics.

Protocol and methods

Male Wistar Han rats (7-8 weeks/200-220 g; Charles River, Germany) were used for all experiments. Steel burrows (32 cm in length and 10 cm in diameter with opening at the front entrance elevated 6 cm from the floor) were filled with 2.5 kg of quartz sand and placed in Plexiglas cages with the open end of the tube positioned so it faced a rear corner of the cage to avoid sand being kicked out of the cage.

The training protocol was in two main phases. The first phase is Social Facilitation (SF) in which rats are placed in the burrowing set-up in pairs on two consecutive days and the amount of sand burrowed is measured. If a pair burrows less than 1500 g of sand on the first day it is swapped with a pair which has burrowed greater than 1500 g. The purpose of the SF phase is to ensure the rats learn the behaviour and the natural tendency of rats to imitate each other's behaviour means that this phase strengthens the burrowing behaviour.

The second phase is individual training where rats are allowed to burrow for 30 minutes a day over three days and the amount of sand burrowed is averaged to get a baseline value for each animal. Any rat burrowing on average less than 1000 g of sand or has a standard deviation of greater than 450 g over the three days is excluded, which usually accounts for less than 2% of rats in a study. The rats are then assigned to treatment groups based on their baseline burrowing values so that each group has a comparable baseline burrowing average.

To induce an OA-like state in rats MIA dissolved in 0.9% physiological saline is injected into the knee joint through the patellar ligament. In our first characterization studies we compared burrowing deficits in rats injected unilaterally with MIA to those injected bilaterally with MIA 3 days after injection (3 mg/knee in 50 μL). Secondly, we investigated the concentration-responsiveness of the deficits at 3 days by injecting 0.3, 1 and 3 mg/knee bilaterally in three groups of rats. The final characterization experiment was to study the temporal pattern of these deficits at 3, 14, 21 and 28 day time points post MIA injection.

For pharmacology studies the efficacy of a range of analgesics in reversing deficits in burrowing behaviour were investigated 3 days after MIA injection. The drugs tested were: ibuprofen (3, 10 and 30 mg/kg, 2 hour pre-treatment time); celecoxib (3, 10 and 30 mg/kg, 2 hour pre-treatment time); morphine (0.3, 0.6, 1 and 3 mg/kg, 1 hour pre-treatment time); anti- nerve growth factor (NGF) monoclonal antibody (mAb) (9 mg/kg, 24 hour pre-treatment time); the Fatty Acid Amide Hydrolase (FAAH) inhibitor PF-04457845 (10, 30 and 100 mg/kg, 2 hour pre-treatment time). Ibuprofen, celecoxib and PF-04457845 were administered periorally (p.o.) suspended in 0.5% Natrosol and 0.1% Tween-80 (9:1 ratio) as vehicle. Morphine and the anti-NGF mAb were administered subcutaneously (s.c.) with saline and phosphate buffered saline as vehicle, respectively. Additionally, in pharmacological studies the burrowing performance of rats was measured 1 day after MIA injection and any rat burrowing greater than 1000 g was excluded.

Results

The first experiments were to characterize burrowing deficits in the MIA model. We found that only bilateral MIA injections produced a large deficit (60% reduction in burrowing compared to sham group) in burrowing performance compared to sham injected animals (50 μL of 0.9% saline into each knee joint) and that 3 mg/knee was the concentration that produced the most robust and reliable deficits with the least variance (811 ± 122g). A time-course experiment showed a large deficit in burrowing performance at 3 days after injection, which resolved to baseline at day 14 and reappeared at 21 and 28 days after MIA injection. The 3 day time point was chosen for pharmacology as it produced a robust deficit large enough to provide an assay window for reversal by analgesics. It was found that burrowing deficits at the 3 day time point were reversible by ibuprofen (10 and 30 mg/kg), celecoxib (3, 10 and 30 mg/kg) and the anti-NGF mAb (9 mg/kg). Morphine did not significantly reverse burrowing deficits at any dose tested and 3 mg/kg caused a significant reduction in burrowing in sham animals, presumably due to the sedative side effects of this drug. The FAAH inhibitor PF-04457845 did not reinstate burrowing behaviour at any dose tested.

Conclusions

Pain induced by intra-articular injection of MIA impairs burrowing performance and this is reversible by analgesic compounds. This demonstrates that burrowing behaviour provides an additional read-out for preclinical assessment of pain behaviour in the MIA model of OA. Additionally, burrowing is sensitive to side effects of drugs such as morphine, which induces a deficit in burrowing performance at 3 mg/kg. This means that both analgesic efficacy and side effect profiling can be achieved in one assay, whereas currently side effects must be assessed via separate methods, such as the rotorod performance test.

Ethical Statement

All animal experimental protocols used for these experiments were authorized by the Local Animal Care and Use Committee and carried out according to the local animal care guidelines, AAALAC regulations, and the USDA Animal Welfare Act.

References

[1] Blackburn-Munro, G. (2004). Pain-like behaviours in animals – How human are they? *Trends in Pharmacological Sciences,* ***2,*** 299-305.

[2] Vierck, C.J., Hansson, P.T., Yezierski, R.P. (2008). Clinical and pre-clinical pain assessment: Are we measuring the same thing? *Pain,* ***135***, 7-10.

[3] Andrews, N., Legg, E., Lisak, D., Issop, D., Richardson, D., Harper, S., Huang, W., Burgess, G., Machin, I., Rice, A.S.C. (2012). Spontaneous burrowing behaviour in the rat is reduced by peripheral nerve injury or inflammation associated pain. *European Journal of Pain,* ***16***, 485-495.

[4] Rutten, K., Robens, A., Read, S.J., Christoph, T. (2014). Pharmacological validation of a refined burrowing paradigm for prediction of analgesic efficacy in a rat model of sub-chronic knee joint inflammation. *European Journal of Pain,* ***18***, 213-222.

[5] Pomonis, J.D., Boulet, J.M., Gottshall, S.L., Phillips, S., Sellers, R., Bunton, T., Walker, K. (2005). Development and pharmacological characterization of a rat model of osteoarthritis pain. *Pain,* ***114***, 339-346.

Correlation of Natural Rat Behaviors in Trait Impulsivity

A. Wearn, A. Pekcec, J.R. Nicholson and N. Gorodetskaya

Department of CNS Diseases Research, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss Germany

Background to Impulsivity

Impulsivity is a key behavioral trait varying greatly among both human and animal individuals, arising as a result of structural and functional changes in the brain [1, 2]. It is characterized by individuals having a predisposition towards rapid, unplanned reactions to stimuli without regard to negative consequences or alternative actions. Increased impulsivity is often observed along with other behavioral abnormalities as being associated with a number of psychiatric disorders such as: substance abuse, Binge Eating Disorder and Attention Deficit Hyperactivity Disorder.

Aims

To investigate whether trait impulsivity correlates with other aspects of natural behavior in a preclinical setting we explored the relationships between locomotor activity, burrowing behavior, grooming behavior and impulsive action in the Lister Hooded rat.

Experimental Procedures

Animals used were 96 male Lister Hooded rats (Charles River Laboratories, Sulzfeld, Germany), weighing 250-270g at the start of the experiments; age approx. 9 weeks. All experiments have been approved by the local Animal Experimentation Ethics Committee.

We measured impulsive action by means of the 5-Choice Serial Reaction Time Task (5CSRTT). In this task, the rat is placed into an operant chamber (Med Associates Inc., Fairfax, USA) containing 5 apertures and a food receptacle. After initiation of a trial by a nose-poke into the food receptacle followed by 5 s delay the rat is presented with a light flash in one randomly allocated aperture. If the rat nose-pokes into the indicated aperture within the 5 s time interval after the cue then it will receive a reward of one food pellet (TestDiet 1811155 (5TUL) AIN-76A Rodent Tablet 45mg, Sandown Scientific). If the rat nose-pokes into a different chamber than the one indicated or makes no response, this is recorded as an incorrect response or an omission, respectively, and results in a time out period of 5 s during which all lights in the box are extinguished and no reward is given. A nose-poke into any aperture before the stimulus is presented is recorded as a premature response. The premature responses are used as a direct measure of impulsivity (more premature responses = higher impulsivity). Rats were trained in this task according to the standard protocol [3].

We measured burrowing activity using a steel tube, (length ca. 32.3 cm, diameter ca. 10.5 cm and elevation off ground ca. 6 cm) filled with 2 kg of normal quartz sand, placed in an empty plastic cage measuring 600 x 340 x 200 mm, facing one of the rear corners. Rats were habituated to the room for 1 hour before start of experiment. The training protocol is in two main phases. The first phase is social facilitation in which rats are placed in the burrowing set-up in pairs on two consecutive days and the amount of sand burrowed is measured. If a pair burrows less than 1500 g of sand on the first day it is swapped with a pair which has burrowed greater than 1500 g. The purpose of the social facilitation phase is to ensure the rats learn the behavior and the natural tendency of rats to imitate each other's behavior means that this phase strengthens the burrowing behavior. In the next stage of training, rats are placed individually into burrowing cages for 1 hour, and amount burrowed is recorded. This procedure is carried out for at least 3 days and until all burrow greater than ca. 500 g. Once stable performance is seen amongst the group, the average of 4 days burrowing for each rat is used as the final measurement.

Locomotor activity was measured in three categories: distance moved (cm), rearing time (s), and rearing frequency. All three end points were measured automatically by Tru Scan ® Activity Monitors and Tru Scan © 2 software (Coulbourn Instruments, USA). Arenas consist of a 41 cm x 41 cm removable floor tray, and four 40 cm tall plastic walls. Rats were placed individually in the arenas for 30 minutes with the lights on and the 3 parameters were measured. Locomotor activity was measured twice for each rat to ensure the presence of stable performance, with a period of 3 weeks between the two measurements.

In order to measure Grooming Activity, rats were recorded for 14 hours overnight, in the dark using Noldus ® PhenoTyper ® equipment, which consists of a 45 cm x 45 cm floor, plastic walls and a PhenoTyper ® top unit (containing an infrared camera). A food tray and water bottle were attached to one of the walls, so that the rats had ad libitum access to them. Grooming activity was measured in terms of grooming time and grooming frequency (number of separate grooming bouts). A gap of at least 5 s signified separation between grooming bouts and bouts lasting less than 5 s were ignored to differentiate from scratching. Measurements were taken both automatically by EthoVision XT10 software and manually, and compared.

Experiments were carried out according to the following order:

1. First measurement of locomotor activity – 4 days.
2. Training and measurement of burrowing activity – 3 weeks.
3. Second measurement of locomotor activity – 4 days.
4. Recording of grooming videos – 7 nights.
5. 5CSRTT training and screening – 14 weeks.

Spearman Rank Correlations were performed against each natural behavior (burrowing, locomotor activity and grooming) versus % premature responses, the index of impulsivity in the 5CSRTT. 78 rats were compared due to some rats not sufficiently learning the operant task. Extreme high and low impulsive rats were selected from the top and bottom 15%, respectively, of premature responses, yielding 12 rats in each group. The average burrowing, locomotor activity and grooming values were calculated for both the high and low impulsive groups, and compared to each other using two-tailed unpaired t-tests. Statistical significance was set to $p < 0.05$.

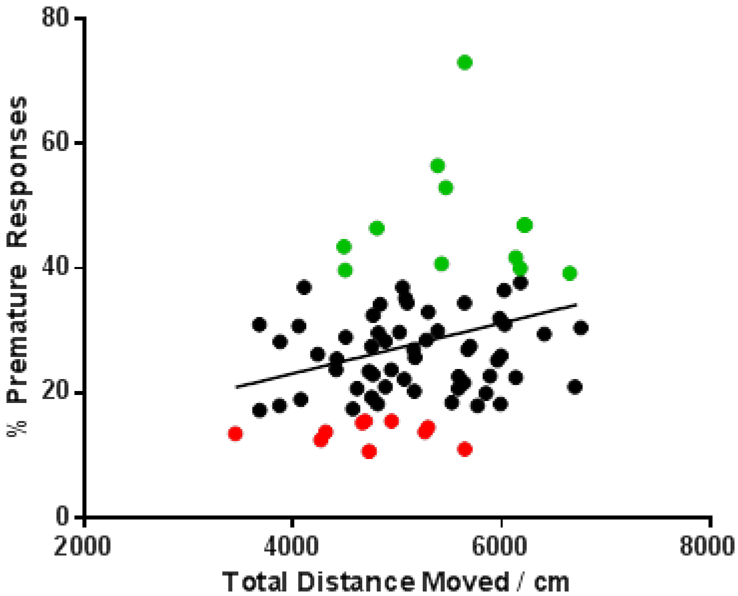


Figure 1. Correlation of first measurement of distance moved and % Premature Responses in 5CSRTT; $^{**}p = 0.0083$; n=78.

Results & Conclusions

We saw no correlation with burrowing activity and premature responses in the 5CSRTT in Lister Hooded rats. The 'Low Impulsive' (LI) group burrowed a mean of 1304±10 g, and the 'High Impulsive' (HI) group burrowed an average of 1250±80 g, between which there is no significant difference. We did see a correlation with distance moved in a novel environment (first measurement of locomotor activity) (***P* = 0.0083; Figure 1) and also rearing frequency in a novel environment (**P* = 0.0103) with premature responses, but not with rearing time, nor with the second measurement of any locomotor activity parameter. The grooming data are currently under analysis. With the data analysed so far, we can conclude that only certain parameters of locomotor activity in a novel environment are related to impulsive action.

References

[1] Jupp, B., Caprioli, D., Saigal, N., Reverte, I., Shrestha, S., Cumming, P., Everitt, B.J., Robbins, T.W., Dalley, J.W. (2013). Dopaminergic and GABA-ergic markers of impulsivity in rats: Evidence for anatomical localisation in ventral striatum and prefrontal cortex. *European Journal of Neuroscience, **37***, 1519-1528.

[2] Caprioli, D., Sawiak, S.J., Merlo, E., Theobald, D.E.H., Spoelder, M., Jupp, B., Voon, V., Carpenter, T.A., Everitt, B.J., Robbins, T.W., Dalley, J.W. (2014). Gamma Aminobutyric Acidergic and Neuronal Structural Markers in the Nucleus Accumbens Core Underlie Trait-like Impulsive Behavior. *Biological Psychiatry, **75***, 115-123.

[3] Bari, A., Dalley, J.W., Robbins, T.W. (2008). The application of the 5-choice serial reaction time task for the assessment of visual attentional processes and impulse control in rats. *Nature Protocols,* **3**, 759-767.

Reanalyzing explorative behavior in the open field test: New insights from a new application that phenotypes mice exploratory behavior

M.M. Tsoory[1], S.Y. Dagan[2], M. Fainzilber[2], N. Panayotis[2]

[1]Department of Veterinary Resources
[2]Department of Biological Chemistry, Weizmann Instiute of Sciences, Rehovot
lmichael.tsoory@weizmann.ac.il

The currently prevailing approach in behavioral phenotyping relies on a defined set of experiments which aim to find abnormalities in a given mouse strain or genotype [1]. Most of the experiments are simple to apply and provide a quantitative, or qualitative, score that can be used as a proxy for various functions such as anxiety, memory capabilities, sensory capacity, motor function and more [2,3].

The open-field (OF) represents one of the most popular tests to assess both locomotion and anxiety in rodents [2,4]. In this test, the animal is placed in an arena and its exploratory behavior is monitored. In its original setups, the assay included an arena with square markings on the floor and exploration was assessed by manually recording the animal's location at designated time frames. Nowadays, automated video tracking systems monitor the animal's location throughout the experiment and supply an output of accurate coordinates [5]. Yet, despite the detailed location these systems provide, most of the analyses include indices that relate to the center and border of the arena (time spent in center; number of visits to center, latency to first visit to the center and total distance traveled). The rationale is that the more a mouse explores the arena and/or the more time it spends in the center of the arena and/or the more frequently it visits the center of the arena, the less anxious it is considered to be.

This dependence on a simplistic distinction between center and border substantially reduces the precision of data analysis and may fail to detect differences between groups of animals of different genotypes or those which underwent different manipulations prior to testing. For example, variations in speed of movement while exploring the arena are rarely considered. In order to allow better phenotyping, we developed a new application for the analysis of mouse behavior in the OF. This application, termed *COLORcation*, relies on analyses of mouse exploration at a user-defined spatial resolution (up to 10 × 10 "tiles") that provides a detailed color heat-map of activity.

The *COLORcation* application utilizes track and rearing files exported from different tracking systems (the presented data was generated from VideoMot2; TSE-Systems) to allow the user to perform a batch analysis of as many groups and mice as desired. The batch analysis parameters may be defined by the user to include different indices, such as number of visits, cumulative duration in a specified location or rearing behavior (if such data is available); furthermore the user may define the number and size of the spatial "tiles" that subdivide the arena.

Figure 1 depicts *COLORcation* "heat-maps" exemplifying the impact of an anxiogenic pharmacological agent (FG7142, 5mg/kg) on the mouse activity in the OF. The "heat-maps" illustrate cumulative duration in each "tile" for each of the compared groups. Such a graphical representation offers an objective illustration of the group activity and may emphasize differences that otherwise would not have been detected in a classic track plot. Further data will be presented exemplifying the impact of restraint stress on the mouse activity in the OF.

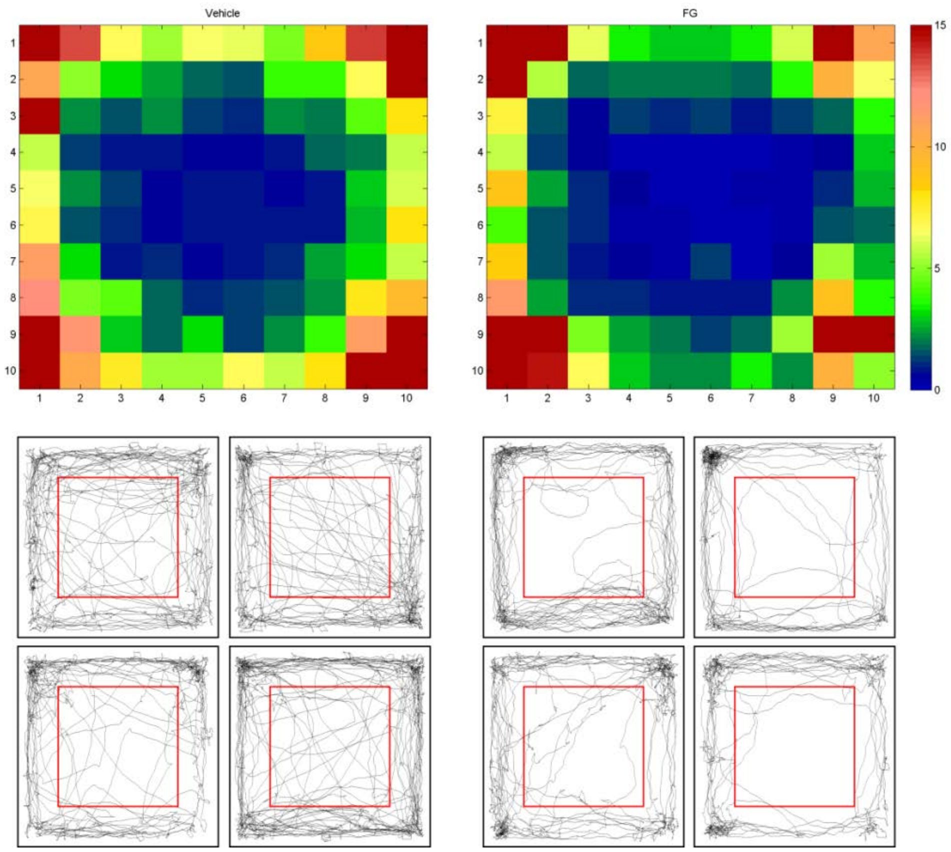


Figure 1. COLORcation heat-map representation. Open-Field activity of WT mice treated with either vehicle (left panels) or an anxiogenic agent (FG7142) (right panels) as depicted in groups heat maps (top panels) comprised of several tracks of individual mice from either group (bottom panels).

Ethical statement

All experimental procedures were approved by the institutional animal care and use committee; approval number 05270813-3.

References

[1] Beckers, J., Wurst, W., and De Angelis, M.H. (2009). Towards better mouse models: enhanced genotypes, systemic phenotyping and envirotype modelling. *Nature Reviews Genetics*, ***10***, 371–380.

[2] Crawley, J.N. (2008). Behavioral phenotyping strategies for mutant mice. *Neuron*, ***57***, 809–818.

[3] Nadler, J.J., Moy, S.S., Dold, G., Trang, D., Simmons, N., Perez, A., Young, N.B., Barbaro, R.P., Piven, J., Magnuson, T.R., et al. (2004). Automated apparatus for quantitation of social approach behaviors in mice. *Genes, Brain, and Behavior*, ***3***, 303–314.

[4] Prut, L., and Belzung, C. (2003). The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *European Journal of Pharmacology*, ***463***, 3–33.

[5] Gould, T.D., Dao, D.T., and Kovacsics, C.E. (2009). Mood and Anxiety Related Phenotypes in Mice. *Neuromethods*, ***63***, 21-41.

Defining Parameters in Automated Quantitative Gait Analysis for Evaluation of Progressive Neurodegeneration in Animal Models of Ataxia and Motor Coordination Impairments

Elisavet I. Kyriakou[1, 2], Jan van de Kieft[1], Huu P. Nguyen[2], Johanneke E. van der Harst[1]

[1] Delta Phenomics, Schaijk, The Netherlands
elisavet.kyriakou@deltaphenomics.com
[2] Institute of Medical Genetics and Applied Genomics, University of Tubingen, 72076 Tubingen, Germany

Introduction

Neurodegenerative disease is a broad term used to describe disorders which have as a common characteristic the progressive death of neurons and the inability to be reproduced or replaced. Depending on the area of the brain where the nerve cell death takes place, neurodegeneration can cause problems with either movement, which in this case the disorders are called "ataxia", or with mental functioning and then it is referred to as "dementia". An increasing number of animal models for several neurodegenerative diseases which exhibit gait disorders have been reported in literature during the past 10 years. However, the lack of an objective and automated method to assess motor coordination and gait abnormalities is making the study of those models difficult, and sometimes the results given are difficult to interpret [1]. CatWalk® is a very promising tool providing an extensive number of gait parameters, however, it has been used so far mainly for modelling pain [2], sciatic nerve injury [3,4] and arthritis [5] but not so much for neurodegenerative diseases that affect locomotion, and more specifically cause ataxia. Therefore, our aim is to validate the system and identify those parameters that can more accurately describe ataxia. To that extend, based on a review of what kind of compounds have been used in literature to induce ataxia, we conducted a pilot study, where alcohol was used as a way to induce temporary ataxia and motor coordination impairments [6, 7].

Materials and methods

Eight adult male Wistar rats (~6 months old, Harlan, The Netherlands) were used, housed under standard conditions with 12:12 h reversed light: dark cycle, temperature 21ºC (±2) housed in Makrolon type IV-S cages (Tecniplast, Italy) with 2 rats/cage. Food and water were available *ad* libitum. Animals gait analysis was performed by using CatWalk XT 10.5 (Noldus Information Technology, Wageningen, The Netherlands). The apparatus consists of an enclosed corridor with a glass plate floor, and a goal-box at the end under which the home-cage of the animals can be put. The runway is illuminated from the ceiling with a red and green light. When the animals' paws have contact with the glass plate the light gets reflected and is captured by a high-speed video camera fixed 60 cm below the corridor. The width of the corridor was approximately set at approximately 8 cm to prevent the animals to turn back and interrupt their straight movement. The automatic detection settings were applied and the green intensity threshold was set to 0.12 and the camera gain to 15.84. The animals were trained for two days during their dark phase, under red light conditions. Before testing, cage-mates were separated for 30min in Makrolon type-III cages with bedding, water and food. This was done to prevent any effects of social interactions and potential hierarchical behaviour on the performance during the test. The animals were habituated to the food rewards and short separation before training and testing. First they were placed on CatWalk to run freely and move across the runway only from left to right. However, based on our previous experience with CatWalk we used an alternative schedule to test and train the animals on the apparatus. That is, each animal was let to have one run on CatWalk, followed by a single run of the next animal. Then the first animal returned on the apparatus for its second run and afterwards the second animal had the second run and so on. This procedure continued until at least 5 straight and with no interruption runs were acquired per animal. Also, the animals were motivated to traverse the corridor by using a food reward after entering their homecage underneath the goal box.

After we established the baseline for all eight animals, two days later the rats were administered with 20% v/v ETOH with a dose of 1.5 g/kg through intraperitoneal injections. They were then tested again on CatWalk 10, 15, 20, 25 and 30 min after the injection. All static and dynamic gait parameters provided from CatWalk software were measured and analysed.

Results

Performance of the rats was determined compared to individual baselines. Data were analysed using the Student's paired t-test, after making sure that all data were normally distributed. Correlations between walking speed and individual gait parameters and between body weight and individual gait parameters were also calculated. Since we tested all parameters CatWalk software produces, we corrected for multiple comparisons by using the false discovery rate test (FDR).

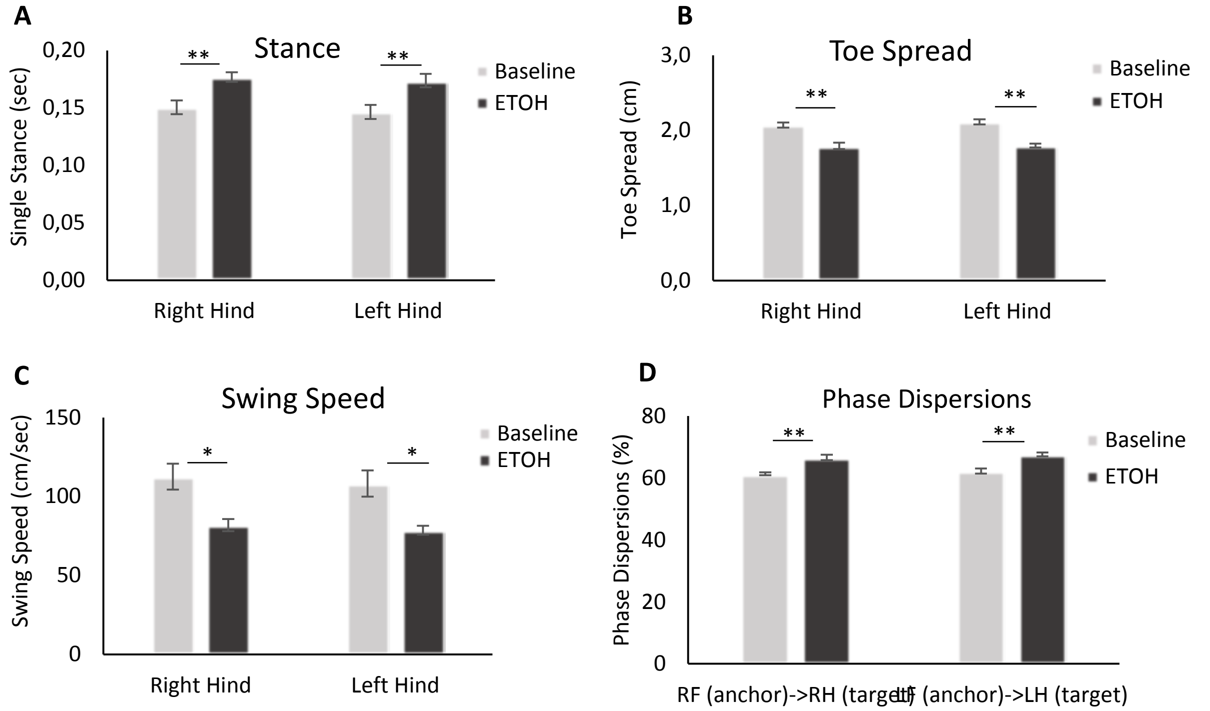


Figure 1. (A) The mean stance duration (s) of ground contact of both hind paws simultaneously and (B) the mean toe spread (cm) before and after alcohol administration. (C) The speed (distance unit/sec) of the paw during swing calculated by the formula: stride length/swing was significantly decreased for both hind paws. (D) The phase dispersions which describes the temporal relationship between placement of the lateral pair of paws within a step cycle and t is used as a measure of inter-paw coordination. The graph shows that the placement of the hind paws (affected) relative to the front paws (unaffected) were significantly delayed compared with the baseline.

After alcohol administration the walking speed (cm/sec), the footfall patterns of the animals and the regularity index (% of regular step patterns) remained intact. While the main reason for the walking speed might be the large variation within the animals, we assume that the intact regularity index might be due to the light motivation of the animal to perform the task in order to receive their reward in the end of each run. However, some other static parameters such as the toe spread (fig. 1B), the print length and the print intensity were significantly affected and more specifically the hind paws. We also observed that all runs acquired under the influence of alcohol were identified as 'compliant' from CatWalk software (maximum speed variation of 50% was selected) suggesting that all treated animals despite their ataxic gait, they had a relatively steady and uniform walking speed compared to their baseline performance. In addition, the dynamic parameters measured by CatWalk, like the stand duration, the swing speed and the phase dispersion (Figure 1A, C and D) showed also a significant effects, which will be presented in a descriptive manner in relation to the potential application as read-outs for future studies associated with various neurodegenerative motor disorders. To that extend, in our future experiments, we aim to use a transgenic rat model for Spinocerebellar Ataxia type 17 [8], and monitor the disease progression over time.

Ethical statement

All experiments reported here were performed with the permission of the Animal Ethics Committee ('Lely-DEC') and in full compliance with the legal requirements of Dutch legislation on laboratory animals.

This study was supported by Marie Curie Initial Network (ITN) of the European Commission's 7th Framework Programme "PhenoRat" (FP7/2012 under grant agreement No. 317259).

References

[1] Wang Y., Bontempi B., Hong S.M., Mehta K., Weinstein P.R., Abrams G.M., Jialing Liu J. (2008). A comprehensive analysis of gait impairment after experimental stroke and the therapeutic effect of environmental enrichment in rats. *Journal of Cerebral Blood Flow & Metabolism.*, ***28***, 1936–1950.

[2] Gabriel A.F., Marcus M.A., Walenkamp G.H., Joosten E.A. (2009). The CatWalk method: assessment of mechanical allodynia in experimental chronic pain. *Behavioural Brain Research,* ***198(2)****,* 477-480.

[3] Bozkurt A., Deumens R., Scheffel J., O'Dey D.M., Weis J., Joosten E.A., Fuhrmann T., Brook G.A., Pallua N. (2008). CatWalk gait analysis in assessment of functional recovery after sciatic nerve injury. *Journal of Neuroscience Methods,* ***173(1)***, 91-98.

[4] Deumens R., Jaken R.J., Marcus M.A., Joosten E.A. (2007). The CatWalk gait analysis in assessment of both dynamic and static gait changes after adult rat sciatic nerve resection. *Journal of Neuroscience Methods,* ***164(1)****,* 120-130.

[5] Fu S.C., Cheuk Y.C., Hung L.K., Chan K.M. (2012) Limb Idleness Index (LII): a novel measurement of pain in a rat model of osteoarthritis. *Osteoathritis and Cartilage,* ***20***, 1409-1416.

[6] Forbes A., Cooze J., Malone C., French V., Weber J.T. (2013). Effects of intermittent binge alcohol exposure on long-term motor function. *Alcohol,* ***47***, 95-102.

[7] Novier A., van Skike C.E., Diaz-Granados J.L., Mittleman G., Matthews D.B. (2013). Acute alcohol produces ataxia and cognitive impairments in aged animals: a comparison between young adult and aged rats. *Alcoholism: Clinical and Experimental Research,* ***37(8)****,* 1317-1314.

[8] Kelp A., Koeppen A.H., Petrasch-Parwez E., Calaminus C., Bauer C., Portal E., Yu-Taeger L., Pichler B., Bauer P., Riess O., Nguyen H.P. (2013). A novel transgenic rat model for spinocerebellar ataxia type 17 recapitulates neuropathological changes and supplies in vivo imaging biomarkers. *Journal of Neuroscience,* ***33(21)****,* 9068-9081.

Tracking behaviour in a large number of experimental units

A. Hettyey[1], G. Nagy[2], Z. Tóth[1], A. Kurali[1], K. Pásztor[1], T. Urszán[2], V. Bókony[1] and G. Herczeg[2]

[1]Lendület Evolutionary Ecology Research Group, Plant Protection Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary
[1] hettyey.attila@agrar.mta.hu
[2]Behavioural Ecology Group, Department of Systematic Zoology and Ecology, Eötvös Loránd University, Budapest, Hungary

Recent technological advances have delivered a virtually endless range of observation tools operational in projects investigating the behaviour of animals. However, experimentalists are often one step ahead in regard to their demands, and it is sometimes difficult to find an appropriate solution for recording movements of focal animals, especially when finances are limited. Cutting-edge research in the study of animal behaviour sometimes requires a large number of video-cameras with rather special specifications. In a study on the quantitative genetic background of animal personality in tadpoles, we had to deploy a video-recording system consisting of 69 cameras, recording swimming behaviour in a total of 800 experimental containers. We describe difficulties of setting up and running such a system on limited resources, and provide simple solutions we found.

We had to position 800 tadpoles kept individually while allowing for recording all of them simultaneously. We used a rack-system with three levels above each other to save lab-space, and arranged tadpole-holding dishpans in groups of twelve to lower the number of cameras needed. We mounted video-cameras on the bottom-side of the shelf above. Recording movements of tadpoles this way, however, required cameras able to take recordings from close-up and exhibiting wide-angle-objectives. While USB-cameras are often a good solution in similar projects, they were not handy in this case because they require external recording and controlling units and we had to use a large number of video-cameras. Between recording-events we had to handle dozens of large video-files and download them relatively quickly, while not disturbing experimental animals too much and not moving the cameras' focus out of place. Consequently, we needed cameras with removable memory cards. We also had to keep costs per unit low and choose thin cameras not to lose precious distance between observed animals and the objective of the cameras. Consequently, handycam-type cameras, surveillance cameras and action-cameras could also not be used. One further challenge we faced was the synchronization of cameras: we had to switch all 69 cameras on and off at the same time and leave the experimental animals undisturbed.

We found that car cameras (also referred to as dashcams) fulfilled all requirements listed above. They record from close-up, exhibit wide-angle-objectives (typically 120°), they come with removable memory cards (which can be upgraded to up to 32 GB), are small in size (starting from ca. 3 cm thickness) and are relatively cheap (starting from ca. 40 € per unit). Furthermore, these cameras are easy to control since they automatically start recording when turned on and stop recording when turned off, making it easy to synchronize them by simply aggregating their supply cables into one single timer. Additionally, an in-built display facilitates aiming of this type of camera before recording. Using these cameras and the setup outlined above, we managed to record movements of 800 tadpoles simultaneously and obtained repeated (15 min × 6 recordings on 2 × 2 consecutive days) good-quality recordings of groups of 12 individuals. We hope our paper will encourage peers to embark on similar projects they had recognized infeasible due to assumedly astronomical costs.

Ethical statement

This research adhered to the legal requirements of the country in which the work was carried out, and to all institutional guidelines. The Közép-Duna-Völgyi KTVF issued the permission to conduct the study (KTVF:10350-2/2012) and the Ethical Commission of the MTA ATK NÖVI approved the investigation in accordance with Good Scientific Practice guidelines and national legislation.

Combining psycho-physiological and social signal processing methods in health communication: Acute stress detection

Kim Groeneveld[1], Marie Postma-Nilsenová[1], Erik Holt[2], Lena Heyn[3], Arnstein Finset[2]

[1]TiCC, Tilburg University
[2]University of Oslo
[3]Lovisenberg Diaconal College

Social signal processing (SSP) offers a range of methods for the semi-automatic measurement of facial, vocal, and bodily expressions. The results of these measurements may be used to analyze the communication between doctors and patients and to provide feedback to doctors in a training setting. The main challenge for SSP methods is to deal with the inherent ambiguity of human expressions. For instance, the experience of acute stress in a patient communicating with her doctor can be reflected in vocal parameters such as fundamental frequency (F0), jitter, shimmer or speaking rate, due to the effect of increased heart rate and bronchodilation (Giddens, Barron, Byrd-Craven, Clark, & Scott Winter, 2013; Orlikoff & Baken, 1989). However, the same parameters can be used as signals of emotional states or even reflect speakers' personality because in general, there is no direct mapping between a cue and its interpretation. Depending on the context and the speaker, vocal stress expression may be highly individual as well (Giddens et al., 2013;[7]. A possible solution to the mapping problem may be the combination of behavioral cues with traditional psychophysiological data. Past studies have shown that bodily reactions accompanying acute stress, such as increase in heart rate, rapid blood flow, activation of sweat glands, and increase in the respiration rate [7] can be measured using modern technology sensors. Among these, the most commonly used method in research on medical interactions involves galvanic skin response measurements, which assess sweating of the skin [1]. In our pilot study involving the analysis of a recorded doctor-patient interaction, we first explored the relation between vocal data and skin conductance measurements. Interestingly, as reported by Kurniawan, Maslov, and Pechenizkiy [7], in some settings, the vocal parameters might be more successful predictors of acute stress than direct physiological measures which differ from person to person, with variances in age, gender, ethnicity, and hormonal cycle. Given that vocal analyses and skin conductance measurements are typically conducted independently of each other, we focused on finding a link between the two types of cues. For this investigation, we used a subset of the data collected by by Heyn, Ruland and Finset [5] and Heyn, Finset, Eide and Ruland [6] with available galvanic skin response measurements. Their corpus consists of conversations of adult patients admitted for initiation of, or continuation of treatment/follow up for leukemia, lymphoma, multiple myeloma, or testicular cancer. In a comparison of short fragments of speech before, during and after a galvanic response episode (T1, T2, and T3), a negative correlation between the slope and SD of fundamental frequency and the galvanic measurements has been found. We also observed a significant increase in Relative Average Perturbation (RAP), a measure of jitter in the voice of the patient, in the fragments of speech selected from T1/T2 and T3. In the second stage of the study, we are currently comparing the vocal and psycho-physiological measurements in a larger sample, using the raw galvanic skin response data.

References

[1] Bakker, J., Pechenizkiy, M., & Sidorova, N. (2011, December). What's your current stress level? Detection of stress patterns from GSR sensor data. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (pp. 573-580). IEEE.

[2] Cohen, S., Kessler, R. C., & Gordon, L. U. (1995). Strategies for measuring stress in studies of psychiatric and physical disorders. *Measuring stress: A guide for health and social scientists*, 3-26.

[3] Glanz, K., & Schwartz, M. D. (2008). Stress, coping, and health behavior. *Health behavior and health education: theory, research, and practice. 4th edition. San Francisco (CA): John Wiley*, 211-236.

[4] Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions*, ***6***(2), 156-166.

[5] Heyn, L., Ruland, C. M., & Finset, A. (2012). Effects of an interactive tailored patient assessment tool on eliciting and responding to cancer patients' cues and concerns in clinical consultations with physicians and nurses. *Patient education and Counseling*, ***86***(2), 158-165.

[6] Heyn, L., Finset, A., Eide, H., & Ruland, C. M. (2013). Effects of an interactive tailored patient assessment on patient–clinician communication in cancer care. *Psycho-Oncology*, ***22***(1), 89-96.

[7] Kurniawan, H., Maslov, A. V., & Pechenizkiy, M. (2013, June). Stress detection from speech and Galvanic Skin Response signals. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on* (pp. 209-214). IEEE.

Implicit Measurement goes Mobile: The Priming Glasses

Ewald Strasser, Christiane Moser & Manfred Tscheligi

ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{Ewald.Strasser, Christiane.Moser, Manfred.Tscheligi}@sbg.ac.at

Introduction

Besides paper pencil approaches, face recognition and physiological measurements, implicit measurement techniques are a great technique for the measurement of emotions. However, implicit measurement approaches lack in the ability for the continuous assessment of emotions because they are stationary. Our approach to deal with this flaw was the development of the Priming Glasses. They combine the ability for stimulus presentation via see-through augmented reality (AR) glasses and measurement of eye movement via eye-tracking. These features enable us to administer implicit measurement techniques, like the Affective Misattribution Procedure (AMP) [1], which are stationary, in mobile settings and continuously during an interaction. By carrying the presentation screen with them, participants can see the stimuli, which are necessary for the implicit measurement techniques, during the interaction. Conventional input via keyboard is replaced by voice input.

Background

1. Paper Pencil Approaches: Emotion questionnaires like the Mood Adjective Checklist (MACL) [2] are good validated methods for the assessment of emotions. In general these questionnaires are not subject to any general specific errors and distortions which makes them not a priori suspicious to be biased in any direction. However, there are limitations that have to be considered when paper pencil methods are used to assess emotions. The most prominent sources of errors are 1) social desirability and the 2) retrospective appraisal of affective situations.

2. Face recognition: A technique that completely avoids the use of reflexive answers is face recognition. This approach is usually based on Ekman's theory [3] of basic emotions. To recognize this basic emotions systems like the face reader [4] try to detect “action units” (FACS) of the face and deduce through calculation of the relation of these units the most likely emotion. The advantage of this approach is apparent, as it is unobtrusive and continuous it is possible to observe emotions over time without altering the experience. However, this approach has also some major limitations. The face recognition is still not mature enough for 1) reliable and continuous measurement due to serial dependence of FACS, which make tracking difficult [5], and 2) the extensive hardware requirements.

3. Physiological measurement: Although also face recognition deals with physiological signals, we think it is helpful to distinguish them from other physiological measurements like EEG, ECG, EDA or EMG [6]. These measures directly assess electrical signals from body surface and they are, therefore, more obtrusive then face tracking. However, they are also independent of verbal reports and can even detect emotions on a level that is hardly assessable by the perceivers themselves. Further they are less suspicious of a social desirability bias and they offer objective results. The main problem with physiological measurements is that 1) they are very intrusive, 2) their use requires special knowledge and 3) that they are very bulky.

4. Implicit Methods: Arguably also physiological measures could be called implicit, but in this context we refer to measures that involve priming techniques and mental association. A common feature of implicit measurement techniques is the presentation of stimuli and the measurement of the reaction of the participants caused by these stimuli. Implicit or better automatic cognition is assumed to be widely independent from explicit cognition. It was shown that implicit measures can predict behavior that cannot be predicted by paper pencil methods (see [7]). Implicit measures are not biased by social desirability (see [7]) like paper pencil methods and they are less bulky then physiological measures. Some of them are also quite extensively validated and they can therefore be

regarded as more mature than face reading. The only challenge they face is that they are currently not able to continuously measure emotions.

The Priming Glasses

From our perspective implicit methods have a great potential for the measurement of emotions. As mentioned above their only challenge is that they are currently not able to provide a continuous measurement of emotion. The Priming Glasses are a first step towards this continuous measurement. As mentioned above implicit measurements have in common that stimuli are presented and the reaction towards these stimuli is measured. The Priming Glasses basically consist of see-through AR glasses (Epson Moverio BT-100), which can be used to present these stimuli. For the unobstrusive measurement of the cognitive load we also implemented head-mounted eye-tracking (based on the ITU Gazetracker software [8]).

The hardware setup of the Priming Glasses uses a hotmirror (see Figure 1) to block external infrared light and reflect the picture of the eye for the eye-tracking camera (see Figure 2) which is equiped with a cold mirror to block visible light. This setup reduces a lot of noise from other light sources. A picture of the device can be seen in Figure 4. A picture from the eye produced by our eye-tracker can be found in Figure 3.

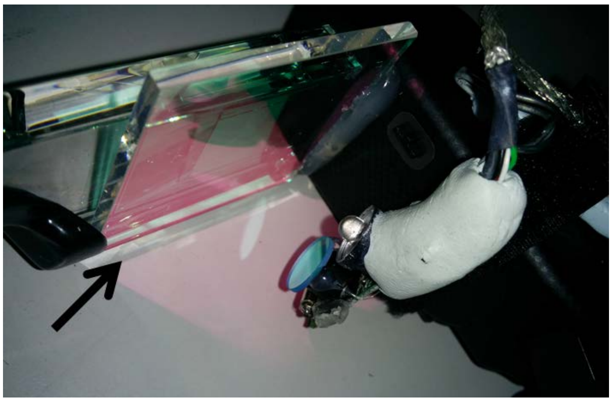


Figure 1. Hotmirror (marked by arrow) reflects image of the eye and blocks external infrared light.

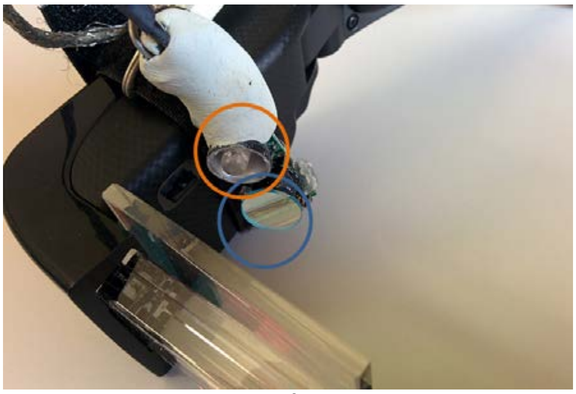


Figure 2. Red circle marks infrared led. Blue circle marks cold mirror in front of the camera lens.

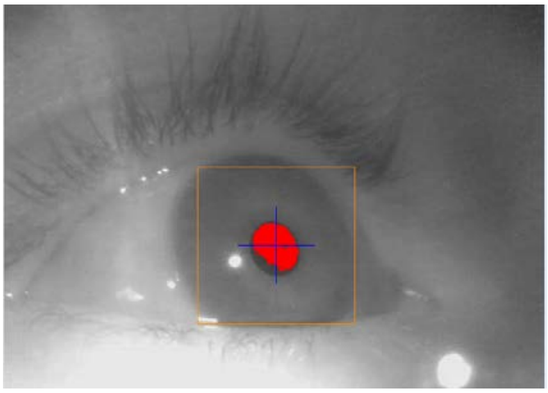


Figure 3. Infrared picture taken with a VGA Camera.

Figure 4. The Priming Glasses in total view.

Future Work

As a first proof of concept we plan to administer an modified AMP over the Priming Glasses. The AMP is a computerized categorization test where participants have to look at a screen and judge whether they like or dislike a Chinese letter. This is of course only the alleged task and the Chinese letter is used as a neutral sign. At the beginning of the task a picture of the objects that should be evaluated is presented for 150ms, this is followed by the presentation of the Chinese letter mentioned before which is presented for 200ms. After that the letter is covered by a mask and participants have to rate if they like the letter or not. However, since the Chinese letter is regarded as neutral the rating should be a result of the emotional appraisal of the previous picture.
Our modification of the AMP involves voice ratings instead of keyboard and we will trigger the AMP trials occasionally during the whole interaction instead of doing it in one continuous session. By aggregating many trails of different participants we will obtain an estimate of the formation of emotions over the time.

References

[1] Payne, B. K., Cheng, C. M., Govorun, O. and Stewart, B. D. (2005) An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of personality and social psychology*, ***89****(3)*, 277.

[2] Nowlis, V. (1965) Research with the Mood Adjective Check List.

[3] Ekman, P. (1971) *Universals and cultural differences in facial expressions of emotion* University of Nebraska Press Lincoln.

[4] Den Uyl, M. and Van Kuilenburg, H. (2005) The FaceReader: Online facial expression recognition. *Proceedings of Measuring Behavior*.

[5] Piątkowska, E. and Martyna, J. (2012) Spontaneous facial expression recognition: automatic aggression detection. *Hybrid Artificial Intelligent Systems*. Springer.

[6] Mikhail, M., El-Ayat, K., El Kaliouby, R., Coan, J. and Allen, J. J. (2010) Emotion detection using noisy EEG data. *Proceedings of the 1st Augmented Human International Conference*. ACM.

[7] Gawronski, B. and Creighton, L. A. (2013) Dual-process theories. *The Oxford handbook of social cognition* 282-312.

[8] San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W. and Hansen, J. P. (2010) Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM.

Low-cost activity trackers for domestic cats and their use in comparing play with free-range behavior

M.S. Sullivan, J. Chapman, S. Watson.

Division of Biology and Conservation Ecology, Manchester Metropolitan University, Manchester, UK

Introduction

Automated data collection on location, and activities of animals has a long history, but the hardware, particularly the batteries, have been so large and heavy as to make them unsuitable for smaller mammals such as domestic cats. Activity and spatial data are of interest for domestic cats for a number of reasons. One is to gather good quality data on their hunting behavior with respect to threats to populations of small animals. Another is to understand the spatial distributions and movements of individual cats over time to explain how domestic cats can often exist at very high densities in urban areas. We can also try to ask more fundamental questions about the relationship between exertion during play behavior and when free-ranging. For example, we could ask whether acceleration during play chases of a toy tends to be higher or lower than when free-ranging. This would provide a further insight into the role of play as physiological 'training' for hunting. The Arduino platform is an open-source and relatively inexpensive microprocessor platform. Because it is open source, a wide range of sensors, communication add-ons and software 'sketches' are being actively developed for a wide range of applications. Our group has been building and testing a variety of wearable sensors and this poster describes how we build and deploy GPS and accelerometer devices.

Methods

For this work we are using Tinyduino which is a modular, small form-factor and lightweight modification of the Arduino platform. We create a stack with an accelerometer, a GPS module, and a microSD card slot for data logging. In addition a good capacity 3v Lipo battery is needed as power is drained quickly by the GPS module and writing to the SD card. Power consumption can be managed in the software by how often the GPS is activated and how often data are read to the SD card. The accelerometer is very low-power but the sampling rate of that too can be managed. The unit is housed in a sturdy, small plastic box and attached to a cat collar or a harness, depending on what the cat is used to wearing. Two cats have been used for prototyping, both of which are of the large breed Maine Coon, which means the kit is particularly small and light compared to body size.

Results and Conclusion

The prototyping indoors for short periods of time has shown that the kit does not interfere with normal activities and is well tolerated. This holds true when actively playing and chasing a toy. Short outdoor trials have confirmed that all the data from GPS and accelerometer are being reliably collected, although there is an issue with calibrating the two signals in time. Early prototyping has allowed us to optimize the power consumption. It has also shown that the range of g required to be monitored is higher than at first thought. Activity and inactivity are clearly visible in the accelerometer data and the algorithm for peak acceleration is still being developed. This small form-factor data logging kit shows definite promise as an activity recording solution for answering questions about domestic cat behaviour. This equipment makes play as training for hunting a tractable question which we are gathering data for at present.

Ethical Statement

The equipment is well tolerated with no distress caused to the cats and it does not interfere with feeding or other activities. The equipment is attached to the collar with a safety release clip. The cats are family pets which are constantly monitored for any signs of discomfort or distress. A repeated-measures design will be used which allows for the smallest number of cats (six) to be used to be able to test for differences in acceleration at $p<0.05$.

Quantification of predator response in groups of fish

Wouter van der Bijl[1], Alexander Kotrschal[2], Niclas Kolm[1]

[1]Department of zoology, university of Stockholm, Stockholm, Sweden
[2]Department of integrative biology and evolution, veterinary university Vienna, Vienna, Austria
[1] wouter.van.der.bijl@zoologi.su.se

Introduction

To understand brain evolution, we need to be able to assess the fitness costs and benefits associated with certain features of the brain. One major component of fitness is survival, which often requires an appropriate response to predators. By studying this predator response in detail we hope to understand in what aspects individuals differ, and to what extent these differences can be attributed to brain morphology.

Escape is the most obvious response to the presence of a predator. Some species, however, perform predator inspections, which is often classified as risk assessment. Here we investigate predator inspection behaviour in the Trinidadian guppy (Poecilia reticulata) towards a model predator, using a complementary approach of video tracking and manual behaviour classification. We used guppies that were artificially selected for large and small relative brain size, differing by 11% after four generations. Because guppies usually occur solitarily and in small groups, we subjected them to the predator model as singletons, pairs and shoals of four.

Methods

We applied continuous top-view recording during the 20 min trials and analysed these videos in two ways. First, by using computer vision software (Ctrax) to track the exact 2D locations of the fish. Second, by manually scoring classic predator inspection behaviour using JWatcher. The created data allows us to exactly quantify the predator response in each trial, e.g. the exact distance to the predator, the speed and direction. The same data can also be used to describe social behaviour, such as the distance between individuals, simultaneous inspections and coordination of movement. Given the nature of the data, for each of these parameters we know both the exact time and location. We can, for example, examine the average speed for each location in the tank. Using time stamps, the computer vision data can further be directly coupled to the manual observations so the location relative to the predator where behaviours were performed can be visualized.

Results and conclusion

Using these techniques we can give the most detailed description of predator inspection behaviour yet, illustrating the power of the complementary use of video tracking. We use this understanding to elucidate the effects of group size, sex and relative brain size on the inspection behaviour of guppies.

Ethical statement

Experiments were performed in accordance with Swedish regulations and approved by the appropriate authority (Stockholm's Norra Försöksdjursnämnd).

Water T-maze as a screening assay for procognitive activity

A.Y.-P. Chen, D.G. Flood, D.A Burnett, L. Leventhal

FORUM Pharmaceuticals, Watertown, MA, USA
achen@forumpharma.com

The water T-maze is an egocentric visual spatial learning task that also incorporates reversal learning. The assay consists of an initial acquisition phase in which the mouse is trained to located a hidden platform located on one side (i.e., right or left). The second phase consists of a reversal in which the platform location is moved to the opposite side. The advantage of this method compared to other water based tasks is that it does not solely rely on visual cues within the test room or a latency endpoint. The current assays utilizes a directional response (vs . spatial) and a correct or incorrect response (vs. time endpoint). Also compared to a standard T-maze task in which the maze is baited the current version does not rely on a food based motivational reward. Previously, the water T-Maze has been used to evaluate cognition in a variety of paradigms: age-related cognitive decline [3, 1], phenotypic profiling of transgenic mice such as Huntington's disease [7, 2] and autism [5] models. Modifications of the assay have been used to address mechanistic questions regarding spatial working memory [6, 8, 4]. However, this assay has not traditionally been used for screening procognitive agents. The current study modifies the protocol described by Guariglia & Chadman (2013) to evaluate the effects of FRM-1, a procognitive compound, on working memory and reversal learning in naïve C57/Bl6 mice.

Methods

Drug Preparation. FRM-1 was dissolved in 0.9% sterile saline and administered subcutaneously in a volume of 5 ml/kg as free base equivalent. Compound was administered 30 min prior to the first test block.

Animals. Male C57/BL6 mice (Charles River, NY) weighing 15-25 g were housed in cages of four on 12h light-to-dark cycle with food and water ad libitum. All studies were performed in accordance with the National Institutes of Health guidelines for the care and use of laboratory animals and were fully approved by the Forum Pharmaceuticals Institutional Animal Care and Use Committee.

Water T-Maze Assay. The maze used was constructed from Plexiglas. Each arm of the T was 35 cm long and 10 cm wide. The T was filled with 23 °C (±1 °C) water to a depth of 21 cm, which was 1 cm above the surface of the platform. Water was made opaque with white non-toxic Tempera paint. The platform was a 5 cm × 10 cm rectangle made from Plexiglas that was made to specially fit into the T-maze and was assigned in a pseudorandom order within groups as either the left or right. At the beginning of each trial, mice were placed in the start area facing one back wall so that no directional bias for swimming was given. Mice were allowed up to one minute to complete each trial. If the platform was not located they were placed on the platform for 10 seconds. For each trial, a mouse was considered to have made a correct choice if it swam directly to the platform. An incorrect choice was recorded if the mouse swam out of the start area in the opposite arm. If the mouse swam out of the start area in the correct direction but returned across the start area before reaching the platform this was recorded as a no choice. Mice were given a total of eight trials per day (2 trials X 4 blocks). The start location was alternated between trials and also between blocks so mice had to learn the correct direction and could not rely on visual cues within the room. Mice were dried with a paper towel and allowed to rest in between blocks for the amount of time it took for all others in the cohort to complete their trials, which was approximately ~30 min. Once the mouse was able to reach daily criterion of 75% correct choices for two consecutive days, it was moved into the reversal learning phase of the experiment. The reversal phase was identical to the acquisition with the exception that the "correct" platform location was now switched to the opposite direction. Data were collected manually by a single blinded observer.

Analysis of Results. For all behavioral analysis GraphPad Prism (GraphPad Software Inc., San Diego, CA) was used for all statistical analysis. Statistical significance between two groups was determined by Mann Whitney t-test. Significant effects across more than two groups were determined by one-way ANOVA or repeated-measure ANOVA where appropriate, followed by Fisher's least significance differences post hoc test.

Results

Acquisition learning. There was no significant difference amongst treatment groups in the number of days needed to reach acquisition criteria ($F(4,72) = 1.5$, $p>0.05$), Fig A). However, a day by day analysis of the data revealed that on day 4, the 3 mg/kg group had significantly greater % of correct choices compared to vehicle ($F(4, 72) = 2.5$, Fig B), with a trend observed in the 1 mg/kg group ($p=0.09$) and no significant effect at either the 0.3 or 10 mg/kg dose groups. One to two animals in each treatment group did not reach criterion during the acquisition phase and were not subjected to reversal learning training.

Reversal learning. During reversal training, there was no significant difference amongst treatment groups over days during reversal learning ($F(4, 58) = 1.0$, $p>0.05$), Fig C). However, a day by day analysis of the data revealed that on day 5, the 1 and 3 mg/kg groups had significantly greater % of correct choices compared to vehicle ($F(4, 57) = 2.0$, Fig D) and no significant effect at either the 0.3 or 10 mg/kg dose groups.

Summary and conclusion

In the current study, FRM-1, a procogntive compound, improved learning in both the acquisition and reversal phases suggesting improved working and reversal (cognitive flexibility) memory, respectively. Overall these data demonstrate that the water T-maze can be used to screen compounds for procognitive activity.

References

[1] Bisaz, R., Boadas-Vaello, P., Genoux, D., Sandi, C. (2013). Age-related cognitive impairments in mice with a conditional ablation of the neural cell adhesion molecule. *Learning & Memory, **20**(4)*, 183-193.

[2] Ciamei, A., Morton, A.J. (2009). Progressive imbalance in the interaction between spatial and procedural memory systems in the R6/2 mouse model of Huntington's disease. *Neurobiology of Learning and Memory, **92**(3)*, 417-428.

[3] Devan, B.D., Pistell, P.J., Duffy, K.B., Kelley-Bell, B., Spangler, E.L., Ingram, D.K. (2013). Phosphodiesterase inhibition facilitates cognitive restoration in rodent models of age-related memory decline. *NeuroRehabilitation, **34**(1)*,101-111.

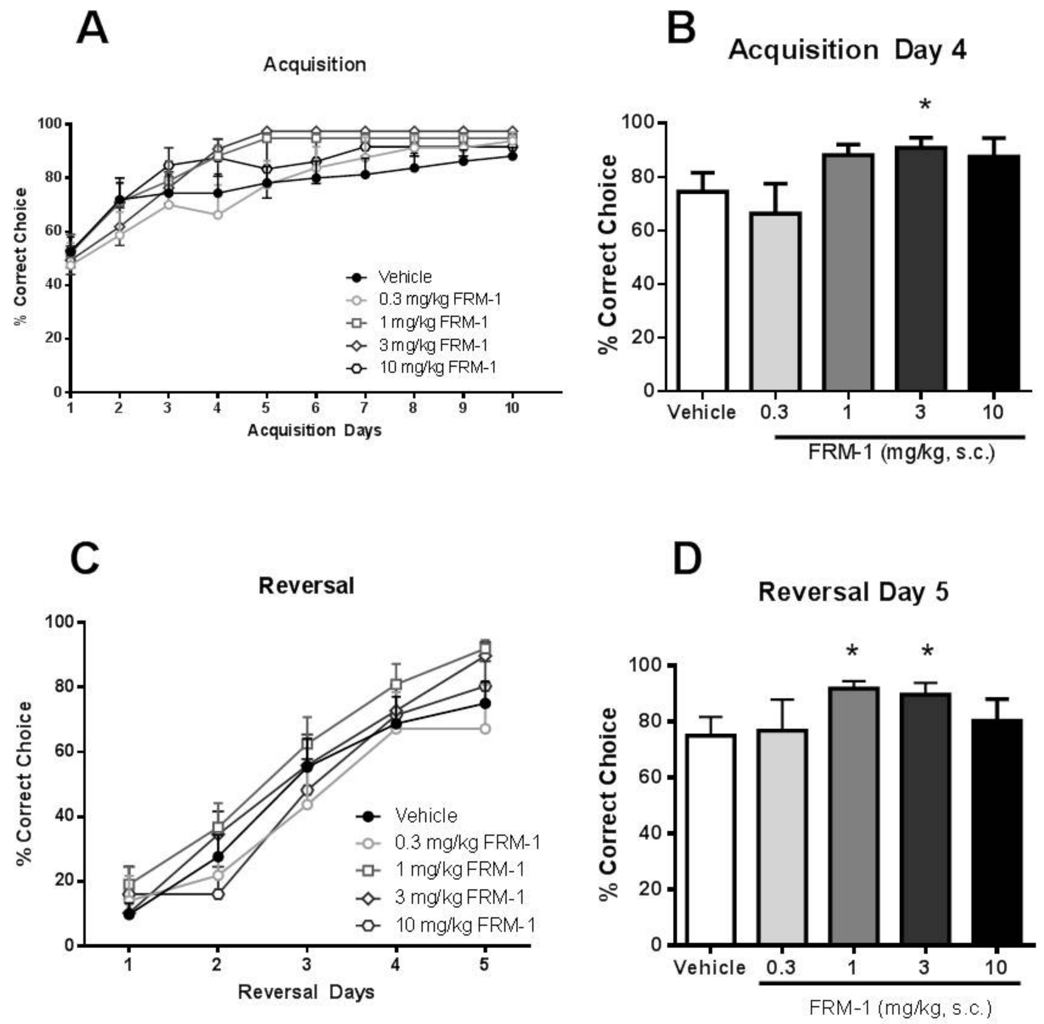
[4] Dwyer, J.A., Ingram, M.L., Snow, A.C., Thorpe, C.M., Martin, G.M., Skinner, D.M. (2013). The effects of bilateral lesions to the dorsal tegmental nucleus on spatial learning in rats. *Behavioral Neuroscience, **127**(6)*, 867-877.

[5] Guariglia S.R.and Chadman K.K. (2013). Water T-maze: a useful assay for determination of repetitive behaviors in mice. *Journal of Neuroscience Methods, **220**(1)**,*** 24-29.

[6] Hawley, W.R., Grissom, E.M., Barratt, H.E., Conrad, T.S., Dohanich, G.P. (2012). The effects of biological sex and gonadal hormones on learning strategy in adult rats. *Physiology & Behavior* **105(4)**, 1014-1020.

[7] Lione L.A., Carter, R.J., Hunt, M.J., Bates, G.P., Morton, A.J., Dunnett, S.B. (1999). Selective Discrimination Learning Impairments in Mice Expressing the Human Huntington's Disease Mutation. *The Journal of Neuroscience, **19**(23)*, 10428–10437.

[8] Maiolia, S., Gangarossaa, G., Locchia, F., Andriolic, A., Bertinic, G., Rimondinia, R. (2012). Excitotoxic lesion of the perirhinal cortex impairs spatial working memory in a delayed-alternation task. *Behavioural Brain Research, **230**(2)*, 349–354.



Figures A-D. Effect of FRM-1 on memory acquisition (A and B) and reversal learning (C and D) in the water T Maze. B, effects of FRM-1 (s.c.) on % correct choice on acquisition day 4 in C57/Bl6 mice (mean + SEM). When compared to vehicle, FRM-1 (3 mg/kg, s.c.) administered 30 min before block 1 significantly improved acquisition learning. Differences from vehicle: *p < 0.05., n=9-20 per treatment. D, Effects of FRM-1 (s.c.) on % correct choice on reversal day 5 in C57/Bl6 mice (mean + SEM). When compared to vehicle, FRM-1 (1 and 3 mg/kg, s.c.) administered 30 min before block 1 improved reversal learning. Differences from vehicle: *p < 0.05., n=7-17 per treatment.

Differential-reinforcement-of-low-rates-of-responding-task performance in a transgenic rat model of Huntington disease

Giuseppe Manfré[1], Huu Phuc Nguyen[1], Valérie Doyère[2], Nicole El Massioui[2]

[1]Department of Medical Genetics, University of Tuebingen, Tuebingen, Germany
[1] giuseppe.manfre19@gmail.com
[2]Centre de Neurosciences Paris-Sud, Unité Mixte de Recherche, Université Paris-Sud, Orsay, France

Huntington's Disease (HD) is an autosomal dominantly inherited progressive disorder which is characterized by psychiatric changes, dementia and motor dysfunction [1]. No effective treatment to influence the onset or the progression of this fatal disease is currently available. The behavioural characterization of a sensitive rodent model for HD is critical, especially in light of the development of therapeutic strategies for HD. In Tuebingen (Germany), Nguyen's group has generated an HD transgenic rat model using a human bacterial artificial chromosome (BAC), which contains the full-length HTT genomic sequence with 97 CAG/CAA repeats and all regulatory elements. BACHD transgenic rats display a robust, early onset and progressive HD-like phenotype including motor deficits and anxiety-related symptoms [2] rendering this model valuable for further phenotype studies.

One of the aspects our research group is interested in is the evaluation of the impulsivity of the BACHD rats. Impulsivity is not a unitary trait, but can be subdivided in two types: impulsive action (IA) is defined as the inability to withhold a prepotent motor response; on the other hand impulsive choice (IC) is the excessive discounting of delayed reinforcement [3]. It is known that basal ganglia are important for the inhibition of unnecessary or inappropriate neural responses: chorea, perseveration during strategy shifts, behavioural disinhibition and impulsivity are the main features that affect HD patients [4]. For these reasons one of the perspectives is the investigation of impulsivity of the BACHD rats. In particular, the DRL (Differential Reinforcement of Low rates of responding task*)* was used to assess impulsive action (IA). This particular task is broadly utilized to evaluate impulsive action, defined as the withhold of an instrumental response for a specific period of time before that response is reinforced [5]. The DRL schedule is an operant schedule that requires the subject to pause for a specified minimum period between responses to obtain a reward. Therefore, efficient DRL performance depends on various components of cognition such as vigilance, self control, short-term memory, waiting ability, and time estimation.

23 male Sprague-Dawley rats (5-months old) were used in this experiment. The w.t. males were supplied from Harlan Laboratories (Italy). The transgenic males (BACHD) were supplied from the Universitätsklinikum Tübingen (Tuebingen, Germany). A 12-h light/dark cycle was maintained during the experiment (light on at 8:00 A.M.). On arrival in the laboratory, rats were group-housed 2 to 4 per cage with wood shavings, were handled on a daily basis and were given access to food and water *ad libitum* until one week before starting the experiment.
One week before starting the experiment the daily food amount was progressively reduced and rats were fed a daily ration to maintain them at 85% of their normal free-feeding weight. Four operant Skinner boxes (31 × 25 × 31 cm) in sound proof ventilated chambers (background noise 65 dB) were controlled with a Graphic State software (Coulborn Instruments, Harvard Apparatus, USA) which controlled programmed task events and data collection. Each chamber was equipped with two left and right transparent walls with transparent back panel and front door. A pellet dispenser for delivery of 45 mg grain-based precision pellets, one 4-cm response lever and a red houselight (lit at the beginning of each session) were located on the right panel.

Training phase began with a 30 minutes-session of magazine training, during which a single food pellet was delivered into the food trough every 60 seconds. On the following day, rats were trained to press a fixed lever in order to receive a single pellet. After achieving a criterion of 50 reinforced lever presses in a 30 min session, rats began the DRL-5. Rats were initially tested on a DRL-5 s schedule for five sessions, during which a lever press only resulted in a single food pellet delivery if at least 5 seconds had elapsed since the previous press. If the rat performed a premature lever press, the 5 seconds time period was reset; thus, rats were only reinforced if they withheld a response for more than 5 seconds. Each DRL session ended either after 60 minutes or 200 reinforcements were earned, whichever came first. Moreover, the red houselight was lit throughout the session. Rats were then tested for ten sessions on a DRL-10 s schedule, during which the response had to be withheld for 10 seconds to obtain reinforcement. Also in this case, each session terminated either after 60 minutes or 200 reinforcements were earned. For each DRL-5 session the frequency of lever pressing during each 1-s time bin (inter-response time; IRT) was recorded. The pattern of responding during DRL-5 was categorized into ‘‘burst,’’ ‘‘premature,’’ and ‘‘timing error responses’’ to include lever presses occurring during 0-1s IRTs, 1-4s IRTs and 4-5s IRTs, respectively. For each DRL-10 session the frequency of lever pressing during each 2-s time bin (IRT) was recorded. In this case ‘‘burst,’’ ‘‘premature,’’ and ‘‘timing error responses’’ took into account lever presses occurring during 0-2s IRTs, 2-8s IRTs and 8-10s IRTs, respectively. Thus, in order to assess performance throughout the sessions of DRL 5 and 10s, measures of instrumental responding, efficiency, genotype, delay and session effect were analyzed.

Analysis of response efficiency in each of the five DRL 5s sessions showed no relevant differences between BACHD and control rats . Two-way ANOVA revealed no genotype effect ($F<1$), a significant main effect of session ($F(4,84)=7.13$, $p<.001$) and a significant interaction between genotype and sessions ($F(4,84)= 4.46$, $p<.005$). On the other hand, response efficiency throughout the ten DRL 10s sessions showed no genotype effect ($F(1,21)=1.35$, ns) but significant effect of session ($F(9,189)=26.82$, $p<.001$). Moreover, no interaction between genotype and session was observed ($F(9,189)=1.49$, ns). In order to assess how the temporal characteristics of lever press responding changed between the two different genotypes, we analyzed interresponse times (IRTs) in three different intervals. There were no significant differences between groups in burst responses (0–1 s IRTs), premature responses (1-4 s IRTs) and timing errors (4-5 s IRTs) of the DRL 5s schedule. Interestingly, burst responses (0-2s IRTs) differences were only evident in the DRL 10s, during which BACHD rats were not able to withold lever pressing. Consequently, the number of responses of transgenic rats was significantly greater than the controls. Two-way ANOVA revealed a significant main effect of genotype ($F(1,21)=4.53$, $p<.05$) and session ($F(9,189)=9.44$, $p<.001$); however, no interaction between genotype and session ($F(9,189)=1.43$, ns) was found. Moreover, there were no differences between genotypes in the premature responses (2-8 s IRTs; FIG. 1F) and timing error (8-10 s IRTs; FIG. 1H) of the DRL 10s schedule. Thus, the main difference between genotypes is evident during the burst (0-2 s IRT) in the DRL 10s schedule, revealing the incapacity of BACHD rats to withold an instrumental response for longer periods of time. Therefore, BACHD rats displayed deficits in behavioral inhibition (i.e. increased responding during non-reinforced periods) that were mainly evident in the DRL 10s schedule. Taken together, these data provide evidence of BACHD rats deficits in the ability to inhibit lever press responding at increasing IRTs , predicting greater impulsive action of this transgenic rat model of HD.

Ethical statement

All experiments were performed in accordance with the recommendations of the European Economic Community (86/609/EEC) and the French National Committee (87/848) for care and use of laboratory animals.

References

[1] Vonsattel J.P., DiFiglia M. (1998). Huntington disease. *Journal of Neuropathology & Experimental Neurology, **57***, 369 –384.

[2] Yu-Taeger L., Petrasch-Parwez E., Osmand A.P., Redensek A., Metzger S., Clemens L.E., Park L., Howland D., Calaminus C., Gu X., Pichler B., Yang X.W, Riess O., Nguyen H.P. (2012). A novel BACHD transgenic rat exhibits characteristic neuropathological features of Huntington disease. *The Journal of Neuroscience, **32(44)***, 15426-15438.

[3] Evenden J.L. (1999). Varieties of impulsivity. *Psychopharmacology (Berl), **146***, 348–361.

[4] Novak M.J., Tabrizi S.J. (2010). Huntington’s disease. *BMJ, **340***, c3109.

[5] Kramer T.J., Rilling M. (1970). Differential reinforcement of low rates: A selective critique. *Psychological Bulletin, **74(4)***, 225-254.

Body weight gain is impaired in rats and mice during automated homecage system observations

L.E. Clemens, D.J. Franke, J.C.D. Magg, E.K.H. Jansson, H.P. Nguyen

Institute of Medical Genetics and Applied Genomics, and Centre for Rare Diseases, University of Tuebingen, Tuebingen, Germany
Laura.Clemens@med.uni-tuebingen.de

Introduction

Detailed behavioral characterization is an important part of establishing novel animal models in a variety of research fields. Automated homecage observation has been promoted to provide better measurements of behavioral parameters than classical tests [1], and several systems for computer-based acquisition of homecage activities of rats and mice as well as analysis software are currently in use [2–7]. These systems offer a standardized testing environment, while also allowing standardized customization to run specific protocols. As the measurements are highly computerized, behavioral data is gathered objectively. Moreover, the systems are able to measure a broad spectrum of behaviors including activity, food and water intake as well as cognitive aspects and thus a well-functioning automated homecage system could in theory be used for the complete behavioral characterization of an animal model.

A further aspect of the promotion by both, companies selling the systems [8, 9] and researchers using them [5, 10–11], concerns the increase in research quality by improving animal welfare. As the tests involve less handling than classical tests, they are believed to be less stressful for the animals, and the results are thought to better represent the animals' natural behavior. The latter is further improved by the opportunity of assessing behavioral parameters over longer periods of time, so that novelty-induced and baseline behavior can be properly separated [1, 5]. However, most systems are designed to measure the behavior of individual animals, and thus require social isolation for the time of observation. Individual housing of rats and mice is usually avoided though, as it is considered stressful and alters an animal's behavioral response [12–13]. From this perspective, both welfare and scientific benefits of using automated homecages might be lost.

Behavioral experiments

It should be noted that data discussed here, were collected in multiple studies as part of the behavioral characterization of different mouse and rat models and as such, the studies were not specifically designed for extracting potential stress effects of the automated homecage environment.

All experiments were approved by the commission for animal experiments at the Regierungspraesidium Tuebingen in accordance with the guidelines of the German animal welfare act. The behavioral experiments were carried out by experimenters trained and experienced in laboratory animal research.

Male, wild type rats and mice were kept under controlled environmental conditions (21–23 °C ambient temperature, 55 +/- 10 % humidity and a 12/12 h light/dark cycle) in two different animal facilities. They were housed in social groups of 3–4 animals in cages with wooden bedding and nesting material. Standard chow and tap water were delivered *ad libitum*.

Automated behavioral phenotyping was performed with the three automated homecage systems listed in Table 1. For this purpose, the animals were transferred to an experimental room and placed individually in the testing cages. The behavior of the animals was investigated over a period of 70 h, during which they had *ad libitum* access to food and water, and were left undisturbed except for a short daily visit. Body weight was measured before and after the test. For some cohorts, body weight was additionally recorded on a weekly basis over a period of several months, enabling the comparison of body weight gain before, during and after behavioral testing.

Table 1. Automated homecage systems used for the behavioral observation of rats and mice. The PhenoMaster systems for mice and rats were provided by TSE Systems, Germany, the PhenoTyper for rats by Noldus Information Technolgy, The Netherlands. Their customized setup of the systems' test cages differed considerably from standard home cages.

System	**Provider**	**Species**	**Cage dimensions** length - width - height	**Peculiarities compared to the home cage**
PhenoMaster	TSE	Rat	48 - 37.5 - 20 cm	Air-tight lid, small free-hanging water bottle and food basket, little amount of bedding
PhenoTyper	Noldus	Rat	45 - 45 - 65 cm	Shelter, two water bottles, large feeding area, non-standard bedding
Home cage		*Rat*	*55 - 38 - 24.5 cm*	
PhenoMaster	TSE	Mouse	20.5 - 36.5 - 14 cm	Small free-hanging water bottle and food basket, little amount of bedding
Home cage		*Mouse*	20.5 - 36.5 - 14 cm	

Results

In our laboratory, automated homecage observation is used for the characterization of animal models of various neurodegenerative diseases [e.g., 14-18]. When analyzing data from these studies, we noticed that the body weight gain of test animals dropped during the test sessions in automated homecages in a similar manner as when running classical behavioral tests (see Figure 1). Comparing the weekly body weight gain of wild type Sprague Dawley rats subjected to behavioral testing to rats that were not tested, revealed a significantly altered development characterized by higher fluctuations in weight gain in the tested rats (see Figure 1a). The effect was even more pronounced when the body weight gain before, during and after behavioral testing was calculated for each rat individually (see Figure 1b), as the exact time of behavioral testing varied between individual rats by up to 12 days.

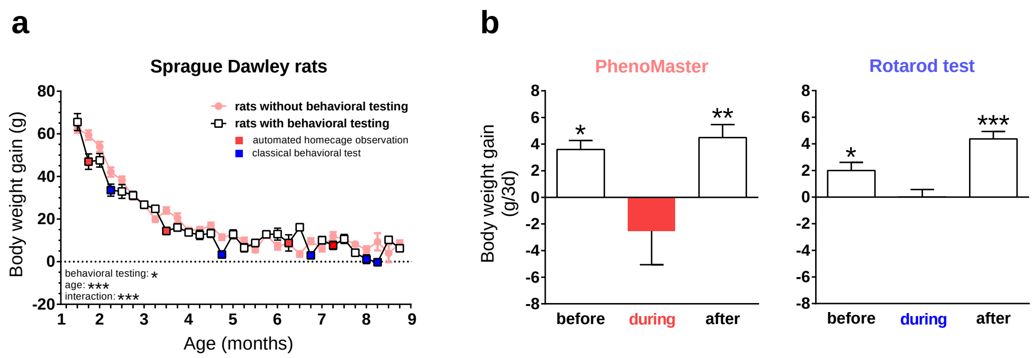


Figure 1. Influence of behavioral testing on body weight gain in Sprague Dawley rats. (a) Body weight gain over 3 days was calculated from weekly measurements for a behavioral test group and a group of rats not subjected to behavioral testing (n = 15 per group). Behavioral test events included observations in the automated PhenoMaster system and classical Rotarod and swim tests. Effects of behavioral testing are likely weakened, since the exact time of behavioral testing varied between individual rats by up to 12 days. Data sets derive from two spatially and temporally separate cohorts of rats. Statistical test results displayed in the graph derive from repeated measurements two-way ANOVA. (b) Body weight gain during the 3 days of PhenoMaster observation and during Rotarod test at the indicated ages are compared to the average body weight gain over 3 days during the week prior and post testing. Data derive from the behavioral test group displayed in Figure 1a and was calculated for each rat individually in order to reveal a more exact measurement. Statistical significance was determined using paired one-tailed t-tests. * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

As reported before, the reduced body weight gain was likely to be the result of reduced food intake, since we found relative food intake of rats with impaired weight gain during automated homecage observation to be obscured [14]. This further influenced other parameters obtained during testing, specifically locomotor activity and metabolic rate, in a similar manner [14].

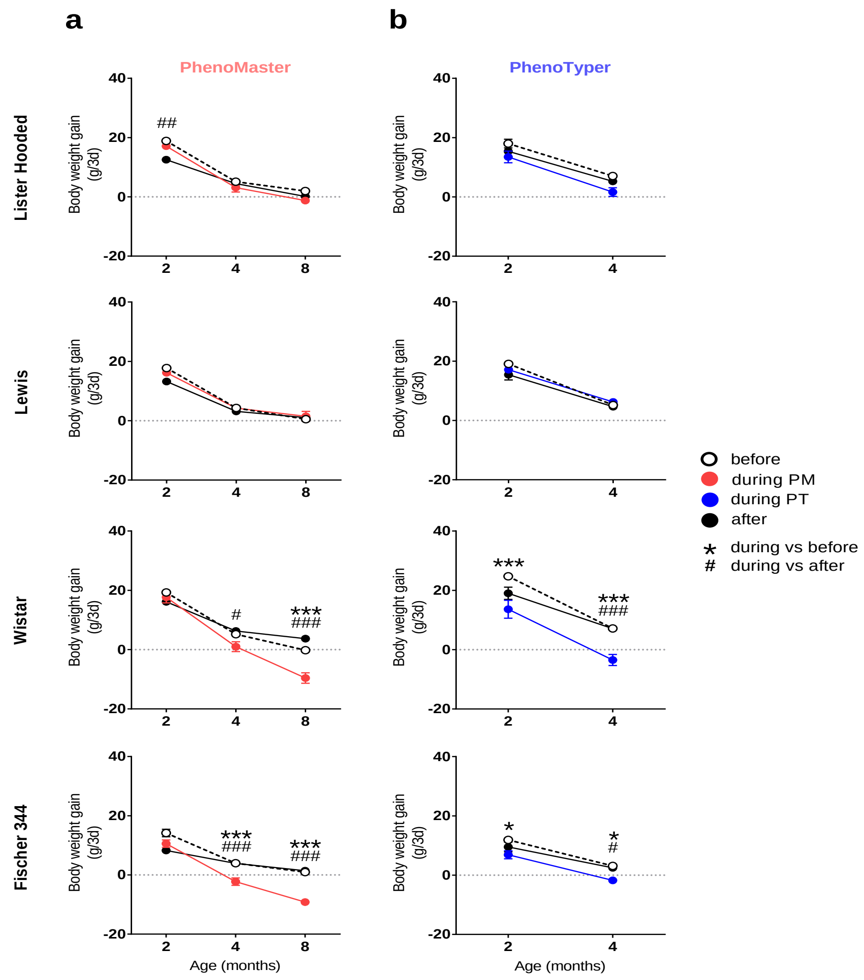


Figure 2. Effect of automated homecage observation on body weight gain in different rat strains. Rats of the four strains Lister Hooded, Lewis, Wistar and Fischer 344 were subjected to automated homecage observation in the PhenoMaster system at 2, 4 and 6 months of age (a) and the PhenoTyper system at 2 and 4 months of age (b). Body weight gain during the 3 days of observation was compared to the average body weight gain over 3 days during the week prior and post testing. Statistical significance was determined using two-way repeated measurements ANOVA and Fisher LSD post test for the indicated comparisons; */# = $P < 0.05$, **/## = $P < 0.01$, ***/### = $P < 0.001$.

The effect of automated behavioral observation on body weight gain increased with age and/or re-exposure to the test cages (see Figure 2a and Figure 3). Furthermore, while younger rats showed a reduced body weight gain during automated homecage observation, older rats experienced explicit weight loss (Figure 2a, Figure 3a). This body weight loss was found as well in aged mice exposed to a mouse PhenoMaster system (Figure 3b).

Discussion

Data presented here, reveal a common impairment of body weight gain during automated homecage observation. This indicates that animals experience some form of stress while being housed in the homecage systems, which is supported by the finding that rat strains regarded as less anxious (i.e. Lister hooded and Lewis rats), did not show the impaired weight gain. The phenomenon, however, seems to be of general importance, as it was found in three out five rat strains investigated, as well as in mice.

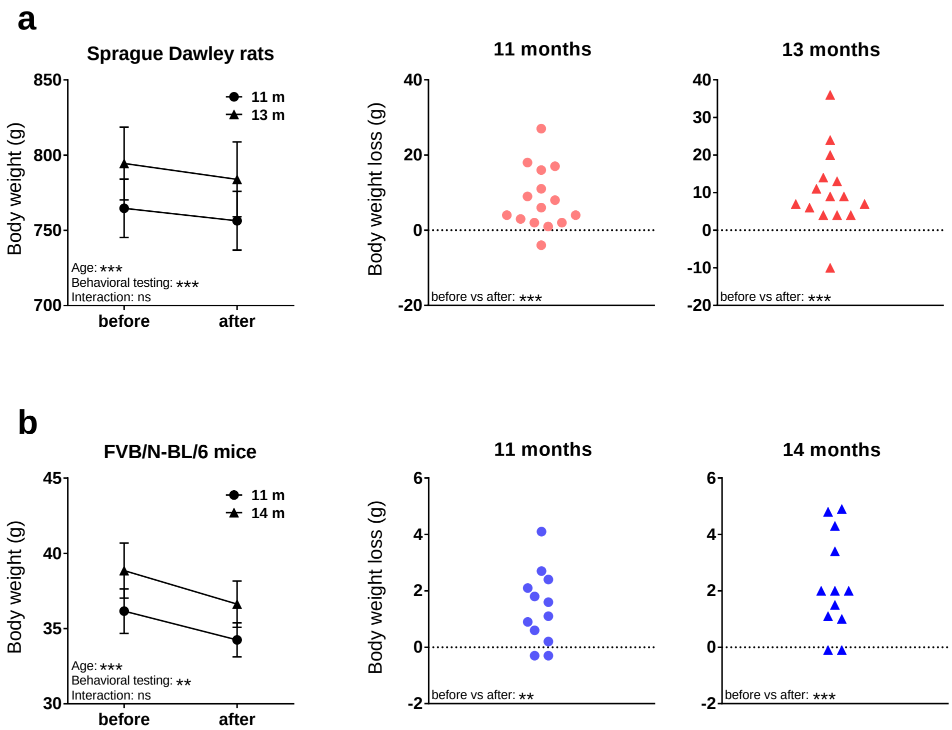


Figure 3. Body weight loss in aged rats and mice during automated home cage observation. Sprague Dawley rats were subjected to automated homecage observation in the rat PhenoMaster at 11 and 13 months of age (a), and FVB/N-BL/6 cross-bred mice were tested in the mouse PhenoMaster at 11 and 14 months of age (b). Body weight was measured before and after 3 days of observation. Statistical test results displayed in the graph on the left side derive from repeated measurements two-way ANOVA. Statistical significance for body weight loss at the two ages (graphs on the right side) was analyzed using paired one-tailed t-tests; * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

Since the impairment occurred in systems from two different manufacturers and in three considerably different setups, it appears unlikely that changes in cage properties have caused the anxiogenic effect. The impaired body weight gain is more likely the result of isolating the animals during the time of observation, as this was the common requirement for behavioral observations in these systems. Similar issues have been discussed for metabolic cages [20–21], and should be addressed as well for automated homecage systems.

It is important to note that the changes in body weight were apparently accompanied by changes in other parameters [14]. Thus, such results need to be interpreted carefully. A more thorough investigation of the anxiogenic factors of automated homecage systems is needed in order to elucidate optimal protocols.

References

[1] Kas, M.J., van Ree, J.M. (2004). Dissecting complex behaviours in the post-genomic era. *Trends Neurosci*., ***27***, 366-369.

[2] Goulding, E.H., Schenk, A.K., Juneja, P., MacKay, A.W., Wade, J.M., Tecott, L.H. (2008). A robust automated system elucidates mouse home cage behavioral structure. *Proc Natl Acad Sci USA*, ***30***, 20575-20582.

[3] Hübener, J., Vauti, F., Funke, C., Wolburg, H., Ye, Y., Schmidt, T., Wolburg-Buchholz, K., Schmitt, I., Gardyan, A., Driessen, S., Arnold, H, Nguyen H.P., Riess, O. (2011). N-terminal ataxin-3 causes neurological symptoms with inclusions, endoplasmic reticulum stress and ribosomal dislocation. *Brain*, ***134,*** 1925-1942.

[4] Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. *Nat Commun*., ***7***, 1-68.

[5] De Visser, L., Van den Bos, R., Kuurman, W.W., Kas, M.J.H., Spruijt, B.M. (2006). Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav*, ***5***, 458-466.

[6] Voikar, V., Colacicco, G., Gruber, O., Vannoni, E., Lipp, H.-P., Wolfer, D.P. (2010). Conditioned response suppression in the IntelliCage: assessment of mouse strain differences and effects of hippocampal and striatal lesions on acquisition and retention of memory. *Behav Brain Res*, ***213***, 304-312.

[7] Zarringhalam, K., Ka, M., Kook, Y.-H., Terranova, J.I., Suh, Y., King, O.D., Um, M. (2012). An open system for automatic home-cage behavioral analysis and its application to male and female mouse models of Huntington's disease. *Behav Brain Res*, ***229***, 216-225.

[8] Homepage of TSE Systems, PhenoMaster/Labmaster, <http://www.tse-systems.com/products/behavior/home-cage/phenomaster/index.htm> Accessed 29 May 2014.

[9] Homepage of Noldus Information Technology' http://www.noldus.com/animal-behavior-research/solutions/research-small-lab-animals/home-cage-monitoring-system. Accessed 29 May 2014.

[10] Lipp, H., Litvin, O., Galsworthy, M., Vyssotski, D.L., Vyssotski, A.L., Zinn, P., Rau, A.E. (2005). Automated behavioral analysis of mice using INTELLICAGE: inter-laboratory comparisons and validation with exploratory behavior and spatial learning. *Proceedings of Measuring Behavior 2005. 5th International Conference on Methods and Techniques in Behavioral Research* (Wageningen, 30 August –2 September 2005), 66-69.

[11] Tecott, L.H., Nestler, E.J. (2004). Neurobehavioral assessment in the information age. *Nature Neurosci*, ***7***, 462-466.

[12] Hall, F.S. (1998). Social deprivation of neonatal, adolescent, and adult rats has distinct neurochemical and behavioral consequences. *Crit Rev Neurobiol*, ***12***, 129-162.

[13] Lukkes, J.L., Watt, M.J., Lowry, C.A.,Forster, G.L. (2009). Consequences of post-weaning social isolation on anxiety behavior and related neural circuits in rodents. *Front Behav Neurosci*, ***3***, 18.

[14] Hübener J., Casadei N., Teismann P., Seeliger M.W., Björkqvist M., von Hörsten S., Riess O., Nguyen H.P. (2012). Automated behavioral phenotyping reveals presymptomatic alterations in a SCA3 genetrap mouse model. *J Genet Genomics*, ***39***, 287-299.

[15] Kelp A., Koeppen A.H., Petrasch-Parwez E., Calaminus C., Bauer C., Portal E., Yu-Taeger L., Pichler B., Bauer P., Riess O., Nguyen H.P. (2013). A novel transgenic rat model for spinocerebellar ataxia type 17 recapitulates neuropathological changes and supplies in vivo imaging biomarkers. *J Neurosci*., ***33***, 9068-9081.

[16] Nuber S., Franck T., Wolburg H., Schumann U., Casadei N., Fischer K., Calaminus C., Pichler B.J., Chanarat S., Teismann P., Schulz J.B., Luft A.R., Tomiuk J., Wilbertz J., Bornemann A., Krüger R., Riess O. (2010). Transgenic overexpression of the alpha-synuclein interacting protein synphilin-1 leads to behavioral and neuropathological alterations in mice. *Neurogenetics*, ***11***, 107-120.

[17] Portal E., Riess O., Nguyen H.P. (2013). Automated home cage assessment shows behavioral changes in a transgenic mouse model of spinocerebellar ataxia type 17. *Behav Brain Res*, ***250***, 157-165.

[18] Yu-Taeger L., Petrasch-Parwez E., Osmand A.P., Redensek A., Metzger S., Clemens L.E., Park L., Howland D., Calaminus C., Gu X., Pichler B., Yang X.W., Riess O., Nguyen H.P. (2012). A novel BACHD transgenic rat exhibits characteristic neuropathological features of Huntington disease. *J Neurosci*., ***32***, 15426-15438.

[19] Clemens, L.E., Jansson, E.K.H., Portal, E., Riess, O., Nguyen, H.P. (2014) A behavioral comparison of the common laboratory rat strains Lister Hooded, Lewis, Fischer 344 and Wistar in an automated homecage system. *Genes Brain Behav*, ***13,*** 305-321.

[20] Gil, M.C., Aguirre, J.A., Lemoine, A.P., Segura, E.T., Barontini, M., Armando, I. (1999). Influence of age on stress responses to metabolic cage housing in rats. *Cell Mol Neurobiol*, ***19***, 625-633.

[21] Kalliokoski, O., Jacobsen, K.R., Darusman, H.S., Henriksen, T., Weimann, A., Poulsen, H.E., Hau, J., Abelson, K.S.P. (2013). Mice do not habituate to metabolism cage housing--a three week study of male BALB/c mice. *PloS One*, ***8***, e58460.

Careful adjustment of food deprivation levels is important when assessing motivation in obese transgenic animals

E.K.H. Jansson, L.E. Clemens, O. Riess, H.P. Nguyen

Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, Germany
Centre for Rare Diseases, University of Tuebingen, Tuebingen, Germany
hoa.nguyen@med.uni-tuebingen.de

Huntington disease (HD) is an autosomal dominantly inherited neurodegenerative disease with a prevalence of 6 per 100 000 in Europe and North America. Development of HD is dependent on a single mutation that results in the extension of the CAG repeat sequence present in the gene for the Huntingtin protein. The major neuropathological hallmark of the disease is a gradual degeneration of the basal ganglia, and HD patients display a range of symptoms that can be grouped into motor, psychiatric, cognitive and metabolic symptoms. There is currently a growing interest for the cognitive symptoms of HD, as they have been found to be present at early stages of the disease, and might thus prove to be of importance when tracking disease progression and treatment effects.

Many tests for cognitive function in rodents are based on food depriving the animals and having them perform certain tasks to retrieve food rewards. For such tests, animals are typically food deprived until they reach a specific body weight. However, the standard protocols might not be entirely suitable for HD models, due to the metabolic disturbances that are often present. For example, animals that express the full-length mutated human protein often show an increased body weight, due to increased amounts of adipose tissue. It is possible that depriving these animals and their wild type (WT) counterparts to the same relative body weight would lead to different hunger levels. Although all behavioral protocols might not be dependent on animals being equally hungry, this is still a preferred situation, as the effect that motivational differences can have on performance is not known for many tests.

Animals' food interest can be assessed by measuring the amount of food consumed during a brief period of free access. In HD research, such tests have been used to match animals' food consumption rates when running operant conditioning tests [1,2] in order to ensure equal hunger and food interest. Thus, we have investigated, whether this food deprivation strategy is suitable for our recently established BACHD rat model of HD. These rats carry a large construct containing the full-length gene for human mutant Huntington, with its endogenous regulatory sequences [3].

First, we concluded that although there was no difference in body weight, BACHD rats carried significantly more adipose tissue compared to WT rats. Further, apart from a brief period of over-eating at young ages, BACHD rats typically ate less food compared to WT rats. We then studied WT and BACHD rats' performance in a progressive ratio test, using different food deprivation strategies. In the progressive ratio test, rats are prompted to lever-push for a small food reward. During a test session, the number of lever pushes required for reward delivery gradually increases, and rats eventually start to lose interest. A common measurement of the rats' motivation is the number of rewards they obtain before taking a break of a specified minimum duration. When BACHD and WT rats were both deprived to 85% of their respective free-feeding body weights, BACHD rats were clearly less motivated than WT rats in the progressive ratio test. In HD research, such a phenotype is of interest, as it could be an indicator of apathy, a common symptom among HD patients. However, BACHD rats also responded with more pronounced drops in motivation when being fed a specified amount of food before the daily test session. In addition, they consumed less food compared to WT rats when given free access for a brief period of time. Both these phenotypes could be indicators of a lower hunger level, which naturally would reduce the rats' interest in lever pushing for a food reward.

To evaluate the use of the aforementioned food deprivation strategy, the food deprivation level of the WT rats was then adjusted so that their food consumption rate was equal to that of the BACHD rats. At that point, BACHD and WT rats showed equal motivation in the progressive ratio test, and also responded equally to the prefeeding tests. Thus, the initial motivational deficit that was present among BACHD rats appeared to be the result of using a suboptimal food deprivation strategy, leading to BACHD rats being less hungry than WT rats. The study highlights the need for careful adjustments of food deprivation levels when working with obese animals, and supports the use of brief food consumption tests for minimizing motivational differences due to differences in hunger levels.

Ethical statement

All experimental procedures were approved by the commission for animal experiments at the Regierungspräsidium Tübingen in accordance with the guidelines of the German animal welfare act.

References

[1] Yu-Taeger, L., Petrasch-Parwez, E., Osmand, A.P., Redensek, A., Metzger, S., Clemens, L.A., Park, L., Howland, D., Calaminus, C., Gu, X., Pichler, B., Yang, W., Riess, O., Nguyen, H.P. (2012). A novel BACHD transgenic rat exhibits characteristic neuropathological features of huntington disease. *Journal of Neuroscience*, ***32***, 15426-15438.

[2] Oakeshott, S., Port, R., Cummins-Sutphen, J., Berger, J., Watson-Johnson, J., Ramboz, S., Paterson, N., Kwak, S., Howland, D., Brunner, D. (2012). A mixed fixed ratio/progressive ratio procedure reveals an apathy phenotype in the BAC HD and the z_Q175 KI mouse models of Huntington's disease. *PLoS Currents*, ***25***, DOI: 10.1371/4f972cffe82c0 .

[3] Oakeshott, S., Farrar, A., Port, R., Cummins-Sutphen, J., Berger, J., Watson-Johnson, J., Ramboz, S., Howland, D., Brunner, D. (2013). Deficits in a simple visual Go/No-go discrimination task in two mouse models of huntington's disease. *PLoS Currents*, ***7***, DOI: 10.1371/currents.hd.fe74c94bdd446a0470f6f905a30b5dd1

Optimization and pharmacological validation of a set shifting procedure for assessing executive function in rats

R. Troudet, A. Michaux, E. Hanon,Y. Lamberty, E. R.Detrait

UCB Biopharma, Biology department, B-1420 Braine-l'Alleud – Belgium
eric.detrait@ucb.com

Introduction

Deficit in executive functions is one of the core component of cognitive impairments in neurological and psychiatric diseases such as Parkinson's disease or schizophrenia [1, 2, 3]. For decades, extra-dimensional set-shifting tests have been used for assessing behavioral flexibility and executive functions both in humans and animals [4, 5, 6]. The extra-dimensional set shifting in a cross-maze was described by Ragozzino et al. 1999 [7] and Floresco et al. 2006 [8] to assess executive functions in the rat.

Strategy set-shifting tasks have been conducted on either a cross-maze or operant chambers. On the maze, rats are initially trained to make a 90° right turn to receive food reinforcement. A visual cue is randomly placed in one of the choice arms on each trial, but do not reliably predict the food location. During the set-shift, the rat is now required to use a visual-cue discrimination strategy, entering the arm with the visual cue, requiring either a right or left turn. Thus, the rat must shift from the old strategy and approach the previously irrelevant cue in order to obtain reinforcement. The capacity of rats to shift strategy is considered as a measure of behavioral flexibility, and is assessed by the number of trials and the number and type of errors made to learn the rule.

The current work aims at optimizing Floresco et al. (2006) set –shifting protocol for standardized drug testing [8]. The following parameters were modulated: caloric restriction, reward preference, length of daily session, definition of turn bias. The new protocol has then been used to assess behavioral flexibility in rats with pre-frontal hypo function induced by sub-chronic phencyclidine (PCP) administration (2mg/kg bid for 7days followed by 7 days wash out). Reversal of PCP-induced deficit have been demonstrated with tolcapone, a brain penetrant COMT inhibitor shown to increase dopamine turnover and function in prefrontal cortex.

Material and method

Male Lister Hooded rats (Harlan, France) were group housed with *ad libitum* access to water and food until beginning of caloric restriction, in a temperature and humidity controlled animal facility (12h : 12h light:dark cycle). Rats were caloric restricted to 48% of their daily food intake in order to reach 85% of their initial body weight after about 2 weeks (Figure 1). All animal procedures were conducted in strict adherence to the European Union Directive 2010/63/EU and were approved by UCB ethical committee. Behavioral experiments were carried out between 9:00 and 16:00, in a sound attenuated and air-regulated experimental room.

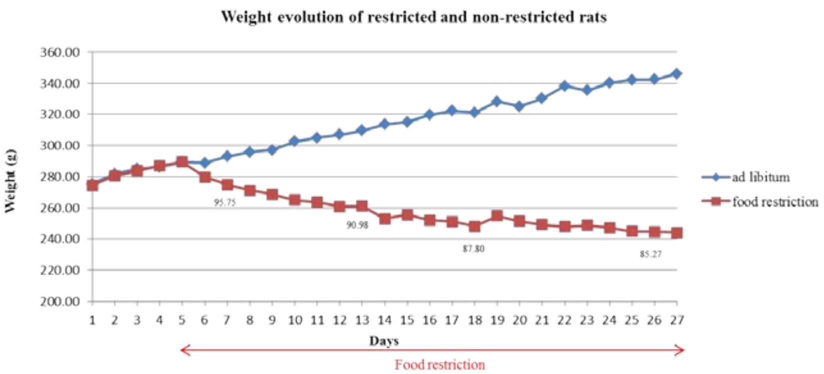


Figure 1. Daily body weight of restricted (n=10, red curve) and ad libitum fed (n=2, blue curve) rats. Numbers under red curve indicated weight, in percentage, compared to initial weight.

Preference for palatable food reward

In an attempt to suppress caloric restriction, 12 *ad libitum* fed rats were given the choice between 6 different palatable food during a 30 min trial repeated for 4 consecutive days. The palatable foods were laying in equal amount (equivalent to 2 sugar pills,~90 mg) on a home cage floor devoid of sawdust: sugar pills (BioServ), chocolate cereal (Kellogg's coco pops), chocolate pearls (Jacques Pearls), crunch peanuts (Delhaize), petal corn (De Halm), wheat cereal (Kellogg's all bran). The sequence of food tasting and the time spent to consume each reward were measured. The first two trials were considered as habituations to new food. Food preference was assessed during the last two trials by the sequence at which rat fully ate each reward. Accordingly, a score from 6 to 1 was attributed to each palatable reward for each rat, 6 being the food first eaten. Data are presented in terms of mean ± sem. Food preference was analyzed with a mixed design ANOVA with the subjects as a random factor and the “palatable food” as a fixed factor.

Cross maze task

The test was preceded by 3 individual habituations to food reward run as the habituations to palatable food preference described above. The maze was a white PVC cross-maze, in which one arm was closed to form a T-maze, the start arm being the foot of the T (Figure 2). A sliding door defined a start box in the bottom of the starting arm. The close arm randomly changed between trials to avoid extra-maze spatial reference cueing leading to food, and one arm was never used as starting arm. Before rule learning, rats were habituated to the maze in a series of 10 successive trials (two 3 min trials followed by eight 1min trials), during which they were allowed free maze exploration, rewards being located at the end of each arm for the 4 last trials. Preferred turn side was defined during these 10 trials as the first turn choice of each trial. Number of left and right turns defined a turn bias when 8 or more first turn were done toward the same side. Habituation sessions were followed by a series of 60 min daily rule learning sessions during which visual cues were randomly placed on the walls of one arm (black-and-white stripes). Each session comprised a series of rule learning trials spaced by 45s stay in the star box and lasting for max 2 min. The first rule leading to reward was egocentric (e.g. figure 2A: right turn, independently of visual cue location). Rule was considered as learnt after 10 successful consecutive trials and two successful probe tests. Probe test was a trial starting from the arm never used before. The day following completion of first rule learning, a shift in rule was applied, reward location being cued by a visual cue (Figure 2B). Number of errors and number of trials to learn each rule were recorded. Error types during the second rule learning were further analyzed to decipher learning strategy. Succession of errors were gathered by bins of four errors, not considering successful trials. A bin was defined as perseverative errors when rat applied 3 or 4 times the egocentric rules, which indicated that rat did not shift strategy (Figure 2C). Bins were defined as regressive when rats made less than three perseverative errors, and indicated a beginning of a shift in strategy. Finally, bin was defined as never-re-inforced when rats adopted neither the egocentric nor visual rule to choose the turn side. Statistical analysis on number of trials and number of errors to criteria and types of errors was performed with the Mann Whitney test, data are expressed as mean ± sem.

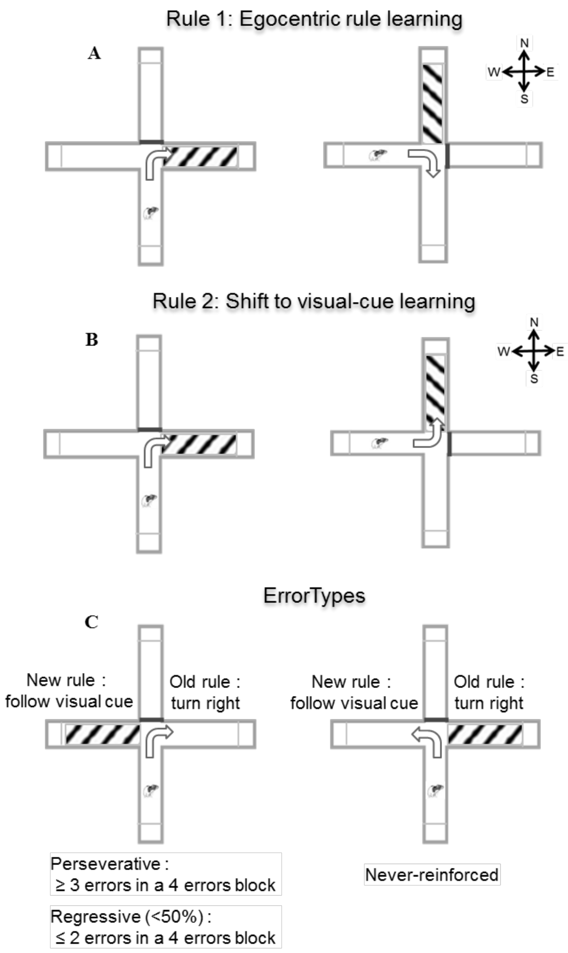


Figure 2. Example of egocentric (A – right turn) and visual (B) rule to be learnt in the set-shifting task. (Arrows represents the correct choice to receive reward). (C) Illustration of error types committed during the shift to rule 2.

Drug administration

Sub-chronic PCP treated rats received intra-peritoneal administration of 2 mg/kg PCP (National Measurement Institute, Australia) twice a day (every 12h) for 7 days in a volume of 5 mL/kg body weight. PCP was dissolved in sterile phosphate-buffered saline (PBS). PCP administration was followed by a 7-day washout period during which rats were daily handled, weighted, and habituated to reward. Control rats received PBS only.

Tolcapone (free base, UCB chemistry department) was suspended in methylcellulose/tween 80/antifoam vehicle (1%/0.1%/0.1%) and intra-peritoneally administered at a dose of 30 mg/kg, 40 minutes prior to each session of rule 2, in a volume of 5mL/kg body weight. Fresh solutions were prepared daily.

Results

The first aim of this work was to optimize Floresco’s protocol to decrease duration and to standardize it for compatibility with an industrial drug testing setting. Habituation sessions in the cross maze were merged to reduce their number from 7 to 1, reward being only located at the end of the arm. Session and trial duration was fixed at 60 min and maximum 2 min, respectively. Identification of turn bias, which is the natural tendency to repeatedly turn right or left, was performed during the 10 trials of habituation to the maze. Protocol optimization also attempted to replace caloric restriction by palatable reward and assessed the robustness of turn bias described in the original protocol.

Analysis of food preference showed that rats developed preference for sweet rewards, which obtained highest preference scores [F(5,55) = 18.3, p<0.0001] (Figure 3). The preferred reward during the second tasting trial was sugar pills (5.17 ±0.30), closely followed by coco pops (4.25 ± 0.22), crunch peanuts (4.25 ± 0.48) and black chocolate pearls (4.00 ± 0.37). All bran and petal corn obtained lowest scores (1.83 ± 0.24 and 1.50 ± 0.20). Data from the first tasting trial were similar to second tasting trial (data not shown). Sugar pills and coco pops were tested as rewards in a set shifting test with *ad libitum* fed rats. 75% of rats failed to fulfill learning criteria for the first rule and stopped searching for reward after about 155.0 ± 30.8 trials. Time to reach reward was much longer than for caloric restricted rats (21.3 s ± 5.9 vs 5.5 s ± 0.6) and it increased with number of trials. Consequently to these poor performances of *ad libitum* fed rats, all subsequent experiments were done with rats under caloric restriction

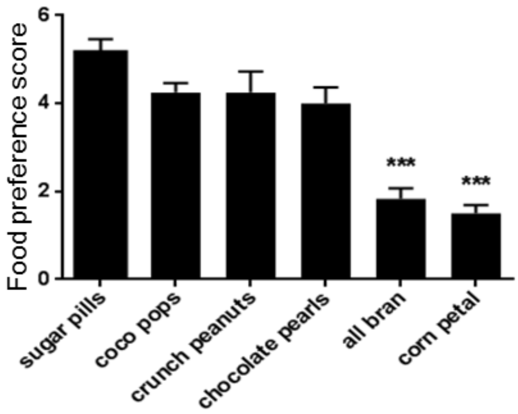


Figure 3. Food preference score evaluated in *ad libidum* fed rats using 6 palatable food rewards. Statistics: mean ± sem, n=12 (***) = p<0.001, compared to sugar pills.

Power analysis was done on turn side to define threshold for turn bias (Figure 4A). In sessions with 7 trials, turn bias toward one side could be defined with a p < 0.05 when rats made 7 turns on the same side (left or right). In sessions with 10 trials, turn bias with a p < 0.05 could be defined when rats made 8 or more turns on the same side. Analysis of natural turn bias in Lister Hooded rats were done on 27 rats using a 10 trial session (Figure 4B). 23 rats made less than 8 out of 10 turns toward the same side indicating that about 85% of rats made random choice for their turn side. Only 4 rats (15%) showed consistent bias toward right or left turn. Those rats were excluded from further testing in order to avoid bias during the egocentric rule learning.

A

# of trials	# of turn toward same side	p value
7	4	0.273
7	5	0.164
7	6	0.055
7	7	0.008
10	7	0.117
10	8	0.044
10	9	0.010
10	10	0.001

B

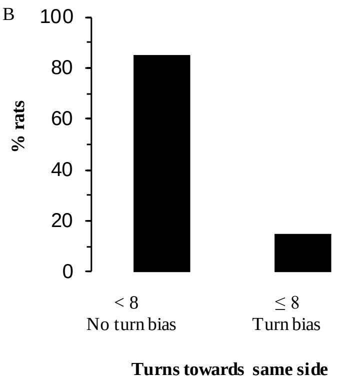


Figure 4. (A) Power analysis defining number of turn toward the same side required to define a turn bias in a 7 and 10 trial session. (B) Percentage of rats (n=27) presenting a turn bias defined as at least 8 turns toward the same side out of 10 trials.

Pharmacological validation of the above optimized set shifting procedure was done using 3 treatments groups: rats treated with repeat vehicle (n=10) were compared to rats treated with sub-chronic PCP receiving vehicule (n=13) or 30mg/kg tolcapone (n=13), administered i.p. , 40 min before each session of second rule learning.

Vehicle treated rat learnt the first egocentric rule in 70.7 ± 8.4 trials and made 19.2 ± 2.7 errors. After learning the first rule, vehicle treated rats were able to shift to the rule based on visual cue; they required 91.9 ± 9.7 trials and made 35.5 ± 6.1 errors for shifting.

Sub-chronic PCP-treated rats learnt the first rule as fast as their controls. They required similar number of trials to reach criterion: 85.18 ± 12.46 trials (U=45, p>0.05). The number of errors to reach learning criterion was 31.0 ± 5.9 and was not significantly different than vehicle-treated group (U=36.5, p>0.05). PCP treated rats thus showed normal capacity to learn an egocentric rule. In contrast, PCP-treated rats performed significantly more errors (58.5 ± 7.1; U=22, p<0.05) and required significantly more trials (159.5 ± 10.8; U=12, p<0.01) to shift to the visual cue; indicating an impairment in their capacity to shift from egocentric rule to visual rule.

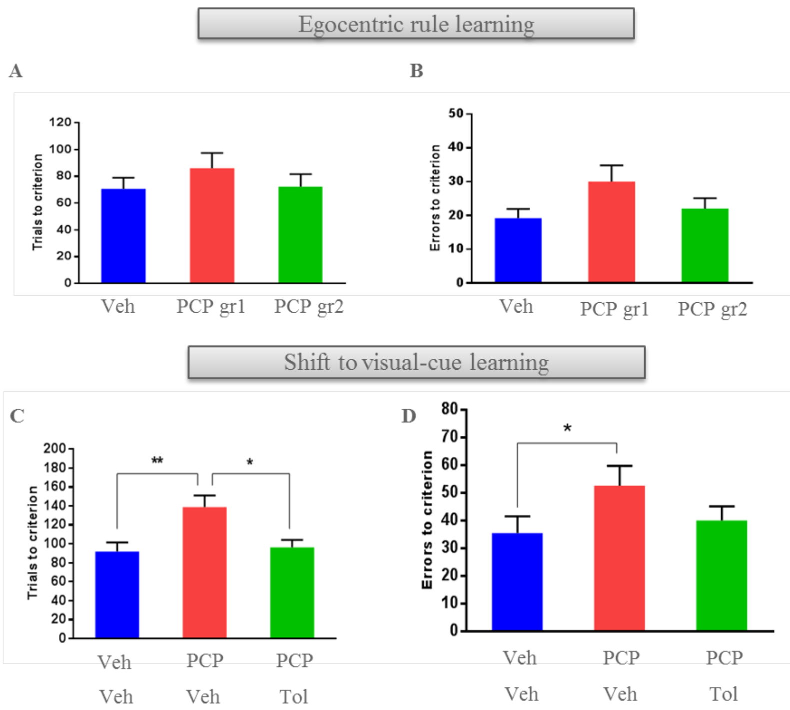


Figure 5. Detrimental effect of sub-chronic PCP treatment on set-shifting and reversal by acute tolcapone administration. (A) Number of trials (A &C) and number of errors (B & D) to reach learning criteria during the egocentric rule 1 (A-B) and visuo-spatial rule 2 (C-D). Blue columns : vehicle only treated rat, red columns: sub-chronic PCP treated rats (2 mg/kg) receiving vehicle before rule 2 learning; green columns: sub-chronic PCP treated rats (2 mg/kg) receiving 30 mg/tolcapone before rule 2 learning. Mean-sem, (*) = p<0.05 and (**) = p<0.01.

Sub-chronic PCP-treated rats receiving 30 mg/kg tolcapone learnt this second rule faster than the PCP-treated rats receiving vehicle. They required 95.9 ± 8.6 trials and made 39.3 ± 5.6 errors to acquire the second rule. The number of trials and number of errors required to learn the second rule was not statistically different than the vehicle treated rats (trials: U=53, p>0.05, and errors: U=50, p>0.05), indicating that 30 mg/kg tolcapone completely alleviated the PCP-induced impairment in set shifting. This is the first time that the brain penetrant COMT inhibitor tolcapone is shown to counteract sub-chronic PCP-induced deficit in extra-dimensional set shifting, a recognized paradigm to evaluate behavioral flexibility.

Detailed analysis of error sub-types demonstrated that all groups mostly made perseverative errors, which is repeatedly applying first rule during second rule learning (Figure 6). Other error types, regressive or never reinforced were marginal. Sub-chronic PCP treated rats made significantly more perseverative errors than other groups (U=32.5, p<0.05 compared to vehicle group ; U=55.5, p=0.14 compared to PCP + Tolcapone group) while regressive or never reinforced errors were similar (p >0.05).

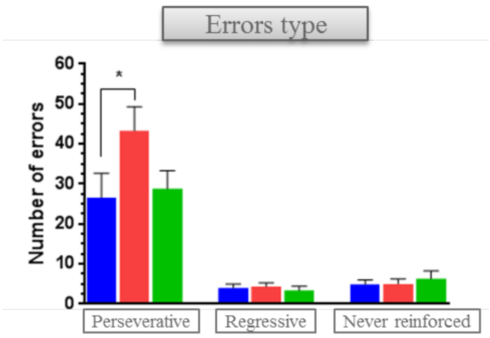


Figure 6. Number of perseverative, regressive and never-reinforced errors committed during the set-shift.

Discussion

The current study aimed at optimizing a procedure assessing executive functions in rodent in order to test new drugs having the potential to counteract cognitive deficits observed in neurological disease such as PD or schizophrenia. The extra-dimensional set shifting task in a cross maze described by Ragozzino et al.(1999) [7] and Floresco et al. (2006) [8] has been modified and pharmacologically validated in a model of subchronic PCP administration. Protocol duration has been significantly reduced by decreasing number of habituations, and by coupling the habituations with identification of turn bias. Criterion to define turn bias has been validated by statistical power analysis. Analysis of turn side in absence of rule learning revealed no particular turn bias in most rats, which led us to introduce an exclusion criterion for the 15% of rats presenting a turn bias.

The attempt to replace caloric restriction by palatable food revealed clear preference for sweet food. Unfortunately, use of preferred food as reward for *ad libitum* fed rats did not seemingly generate enough motivation to complete first rule learning. Consequently, caloric restriction was kept as a mean to motivate rats for learning the task. Further testing would be required to assess whether coupling preferred food either with very low caloric restriction, e.g.; reaching 95-95% initial body weight, or with overnight food deprivation in *ad libitum* fed rats would provide enough motivation for completing the task.

The methodology to compute error types (perseverative vs regressive or never re-enforced) was modified compared to Floresco's protocol to better reflect number of perseverative errors committed during shift to second rule. The current methodology analyzed the succession of errors by bins of 4 errors (excluding successful trials) rather than bins of 4 trials including successful trials. Using such methodology a succession of

2 perseverative errors (P) followed by two successful trials (S) would be classified as regressive bin while the errors are all perseverative. For instance, the sequence of 8 trials [P-P-S-S]-[P-P-S-S] would be counted as 4 regressive errors with Floresco's methodology and as 4 perseveratives errors with the current methodology [PPPP]. As a consequence, Floresco's methodology underestimated bins of perseverative errors (table 1). The current methodology better reflects the increase in perseverative errors committed by PCP-treated rats while Floresco 's methodology did not. In contrast, Floresco methodology is more sensitive to detect the appearance of regressive error bins, which is the time when the animal start shifting strategy.

Table 1. Comparison of methodologies for classification of errors type between protocol described by Ragozzino et al. (1999) and Floresco et al. (2006), and the optimized protocol.

Treatment	**Perserverative**		**Regressive**		**Never-reinforced**	
Methodologies	*Troudet*	*Floresco*	*Troudet*	*Floresco*	*Troudet*	*Floresco*
Veh	26.2	7.5	4	21.3	4.9	4.9
PCP + veh	41.4	9.9	4.4	35.7	5.1	5
PCP + tolcapone	26.6	8.9	3.1	20.9	5.8	5.9

The extra-dimentional set-shifting test is a relevant translational test; adapted from tests currently used in clinical studies to assess executive functions in humans, such as the Wisconsin Card Sorting Test (WCST). Impairments of cognitive flexibility have been observed in schizophrenic patients in this test. Repeated administration of moderate doses of PCP (2 to 2.8 mg/kg) in rats led to a significant decrease in release and use of basal dopamine in the prefrontal cortex, which was subsequently correlated with impaired delayed alternation task [9, 10]. In this study, the optimized protocol was validated in the sub-chronic PCP model of memory deficit. PCP disrupted the shift in strategy to reach the reward without affecting first rule learning. PCP treated rats increased number of trial to criterion and made significantly more perseverative errors to shift to the second rule. PCP-induced deficit was alleviated by 30mg/kg tolcapone administration 40 min before each session of second rule learning. Beside protocol validation, these data provided the first evidence for an efficacy of COMT inhibitor on sub-chronic PCP induced set shifting deficit.

References:

[1] Eisenberg DP, Berman KF (2010) Executive function, neural circuitry, and genetic mechanisms in schizophrenia, *Neuropsychopharmacology, **35(1)***, 258-277

[2] Millan MJ, Agid Y, Brüne M, Bullmore ET, Carter CS, Clayton NS, Connor R, Davis S, Deakin B, DeRubeis RJ, Dubois B, Geyer MA, Goodwin GM, Gorwood P, Jay TM, Joëls M, Mansuy IM, Meyer-Lindenberg A, Murphy D, Rolls E, Saletu B, Spedding M, Sweeney J, Whittington M, Young LJ (2010). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy, *Nat Rev Drug Discov, **11(2)***,141-168

[3] Orellana G, Slachevsky A (2013). *Executive Functioning in Schizophrenia, Front Psychiatry., **4***, 35.

[4] Alvarez JA, Emory E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol Rev, **16(1)***,17-42

[5] Jurado MB, Rosselli M (2007). The elusive nature of executive functions: a review of our current understanding, *Neuropsychol Rev, **17(3)***,213-233

[6] Chudasama Y. (2011). Animal models of prefrontal-executive function. Behav Neurosci., ***125(3)***, 327-343

[7] Ragozzino ME, Detrick S, Kesner RP (1999). Involvement of the prelimbic-infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning, *J Neurosci., **19(11)***, 4585-4594

[8] Floresco SB, Magyar O, Ghods-Sharifi S, Vexelman C, Tse MT (2006). Multiple dopamine receptor subtypes in the medial prefrontal cortex of the rat regulate set-shifting. *Neuropsychopharmacology, **31(2)***, 297-309.

[9] Jentsch JD, Roth RH (1999). The neuropsychopharmacology of phencyclidine: from NMDA receptor hypofunction to the dopamine hypothesis of schizophrenia, *Neuropsychopharmacology, **20(3)***,201-25.

[10] Morris BJ, Cochran SM, Pratt JA. (2005) PCP: from pharmacology to modeling schizophrenia, *Curr Opin Pharmacol, **5(1)***, 101-106.

Validation of a virtual reality tool to test consumer response in supermarket settings

Andersen, M.R.[1], Brisson, P.[1], Hald, P.L.[1], Godtfredsen, D.[1], Serafin, S.[1], Mikkelsen B.E.[2]

[1]AAU-Multisentory Lab, Aalborg University, Denmark
[2]FoodScape Lab, AAU-MENU, Aalborg University, Denmark

Background

Normally, when researchers want to do a study on consumers' behavior, the possibilities are limited to:

- Doing a mockup – having a design either recreate a staging from pictures or create an entirely new stage, out of cardboard or similar (a model approach)
- Building the stage – having to help build the actual stage completely and filling it with real objects (a food lab approach)
- Using the actual location – using the actual location as a stage if possible (a living lab approach)

Recently, researchers have been looking for other possibilities. For example, a team from Holland [1] has created a simulation of a supermarket with a traditional computer setup, meaning screen, mouse, and keyboard and virtual shopping is a growing field of research [2] With recent developments and availability of head mounted displays, such as the Oculus Rift, it is possible to create even more realistic simulations, where actual body movements are used to control the interaction with the computer instead of the mice and keyboard. The obvious advantage is that the food environments are developed virtually at no cost, other than time when creating and texturing the models. The aim of this study was to carry out a usability test of the virtual supermarket tool (VST) that simulate a real life check-out experience at the cash register in a supermarket and allows for studies of food consumption and food choice behaviour.

Methods

The development of the virtual supermarket tool (VST) included the following equipment: the Oculus Rift, a Windows Laptop for running the program, 3 Wii Remotes, a custom-made fingerless glove with three infrared diodes and a button for selecting objects and 2 holsters for placing the Wii Remotes on the subjects' legs. The equipment was tested pre-tested among experts at the campus and then taken into a living lab test among shoppers in a real life users in a supermarket setting. The results showed that further testing among "early mover" users was necessary. As a result a total of 30 total participants was recruited from the university student housing Otto Mønsteds Kollegium. They were asked to test the equipment and answer questions categorized under the following themes visuals, mobility, precision, movement, interaction and ease of looking around. The information gathered was stored on a server, which allowed the researchers to collect and import data into a work sheet for further analysis.

Results

The results from the testing so far of the virtual supermarket proved promising, as the participants felt it was both possible and comprehensible to relate to the shopping routine of each individual. A number of weaknesses were found in the usability test. Results showed that it lacked the capability to compare virtual behavior with real behavior seeing since it was not possible to have the test participants consecutively shopping inside a real-life supermarket. While the virtual reality supermarket was indeed viewed positively by most participants, it nevertheless had certain shortcomings. The weaknesses of the test included the fact that augmented glove was too bulky and fragile and that the lighting of the virtual environment would have been better with Unity Pro.

Discussion

The weaknesses identified could be addressed the following way. The hit-box of the shopping basket would need to have its coding altered so as to achieve the enlargement. The excavated circuit of the mouse tended to block the light from the infrared LEDs and a solution to this would be to re-design the layout of the glove's gadgetry to as to swap the current positions of the LEDs and the mouse circuit, thereby allowing the HMD's Wii remote control a clear view of all three LEDs without the need for tilting the real-life hand backwards at all. The fact that the participants often felt the need for walking backwards or sideways needs and wanted to move faster than anticipated to be further studied. The fact that the participants attempted to keep their grip closed upon trying to grab a virtual item could be addressed through a modification in which each real-life step taken by a user to be represented by two or three steps in the digital world. This would result in which participants would have to walk-in-place a lot less, and thereby get around the store with more ease. While the virtual supermarket concept seems well made, the interactive aspects can be taken even further, through the use of more advanced input- and output devices, so as to optimize the virtual experience even more. However the test showed that the potential test subject might be limited to younger and middle-aged participants due to their familiarity with computers. It can be anticipated that through proper design this tool can be applied to a much wider target group than only middle aged users.

References

[1] Waterlander, W., Scarpa, M., Lentz, D., Steenhuis, I.H.M. (2011). The virtual supermarket: An innovative research tool to study consumer food purchasing behaviour. *BMC Public Health, **11**(1),* 589.

[2] Westland, J.C., Au, G.A. (2014) Comparison of Shopping Experiences Across Three Competing Digital Retailing Interfaces. *International Journal of Electronic Commerce, **2**(2),* 57-69.

[3] Lee, K.C., Chung, N. (2008). Empirical analysis of consumer reaction to the virtual reality shopping mall. *Computers in Human Behavior, **24**(1),* 88-104.

Acknowledgments

Scientific Program Committee

- Dr. Andrew Spink (chair, scientific program committee), Noldus Information Technology, Wageningen, The Netherlands
- Prof. Gernot Riedel (conference chair), Department of Biomedical Sciences, School of Medical Sciences, University of Aberdeen, United Kingdom
- Dr. Egon L. van den Broek (conference chair), Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
- Dr. Maurizio Mauri (conference chair), University Institute for Modern Languages (IULM), Milan, Italy
- Prof. Wim Crusio, French National Research Council, Talence, France
- Dr. Lucas Noldus, Noldus Information Technology, Wageningen, The Netherlands
- Dr. Fabio Paternò, Laboratory on Human Interfaces in Information Systems, Institute of Information Science and Technologies, C.N.R., Pisa, Italy
- Prof. Vicenç Quera, Department of Behavioral Science Methods, University of Barcelona, Spain
- Prof. Ilan Golani, Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Israel
- Prof. Eddy van der Zee, Faculty of Mathematics and Natural Sciences, University of Groningen, Groningen, The Netherlands
- Prof. Boris de Ruyter, Philips Research Europe, Eindhoven, The Netherlands / Radboud University, Nijmegen, The Netherlands
- Prof. Jan van Erp, TNO, Soesterberg, The Netherlands / University of Twente, Enschede, The Netherlands

Local Organizing Committee

- Natasja Bogers, chair, local organizing committee
- Dr. Andrew Spink, co-chair, program committee
- Melvin de Bruijn, finances
- Sabine van der Meijde, conference secretariat

Conference Secretariat

- Sabine van der Meijde
- Nathalie Oostendorp
- Nancy van den Dam
- Mariska Walter
- Sidney Hooijer, finances

First author index