Multi-medial Stress Assessment*

L.J.M. Rothkrantz
Department of Information Technology and Systems
T.U.Delft, The Netherlands
L.J.M.Rothkrantz@ewi.tudelft.nl

R.J. van Vark
Department of Information Technology and Systems
T.U.Delft, The Netherlands
R.J.vanVark@ewi.tudelft.nl

D. Datcu
Department of Information Technology and Systems
T.U.Delft, The Netherlands
D.Datcu@ewi.tudelft.nl

Abstract - *A research project on stress assessment is running at Delft University of Technology since 1992. One of the aims of the project is to develop an instrument for automated stress assessment. The underlying system is based on the analysis of facial expressions, voice analysis and the analysis of physiological signals such as heart rate and blood pressure. Analysis of these multi-medial data takes place in parallel and are based on Artificial Intelligence technology. In each of the parallel subsystems, corresponding to sensor, image and sound data, the functionality is split up into a number of layers: filtering and reduction layer, preprocessing layer, processor layer, application layer and output layer. The results of the analysis are combined by a central interpreter, resulting in an overall stress measure. In this paper the stress assessment is used to monitor vigilance levels of car drivers with a focus on voice analysis.*

Keywords: Multimodal interaction, stress assessment, micro sleeps.

1 Introduction

Traffic accidents and other disasters caused by the breakdown at the man-machine interface are significant source of death rate, injuries and environmental damage. Apart from the social aspect of accidental death or injury, the direct and indirect costs of disasters to the Community are enormous. Some of these accidents, caused by downward drifts in vigilance levels sliding into a micro-sleep or cognitive overload of the car driver, are certainly preventable and even a partial success at prevention should have a huge socioeconomic impact. In recent years many advanced driver assistance systems (ADAS) have been developed. By enabling those systems to communicate with each other, internal in the vehicle as well external between vehicles, will result in a major step forward in driver assistance and in road safety. At TUDelft there is a project called "Smart Road" running, aimed at the development of a system that could perform the intelligent tasks of guiding road-users. Large number of small, autonomous sensors installed in the road together form a complete traffic control system. Each sensor unit is a stand-alone intelligent system with configured processors, application specific hardware for detecting traffic and wireless communication. Via a PDA in the car the driver has access to the wireless sensor network. At DECIS-THALES there is a project running on decentralized decision making in a traffic environment. The question is whether it is feasible to control traffic using the concepts of self-organization as embodied in for example "network swarm intelligence". Vehicles are communicating in wireless networks and provide the car driver with information for dynamical routing. The objective of the project is to discuss systems capable of monitoring vigilance levels, predicting the probability of a micro-sleep emergence and warning, as well as training the operators to become aware of this danger. Detection of micro-sleeps with high reliability can only be realized by EEG analysis. But many non-intrusive correlates can be used such as movements of head and eyes, driving behavior and non-verbal content of speech. Based on this measures warning systems can be developed which can be validated by EEG analysis. In this paper we discuss a stress assessment system based on multi-medial data. Such a system can be used as a detection system for the onset of micro-sleeps of car-drivers.

2 Model of multimodal stress assessment

The developed model is modeled analogous to the human brain, as is common in both artificial intelligence and neuro-cognitive sciences. People observe their environment by the use of several sense organs, for instance eyes and ears. On the basis of these observations a meaning can be attached to these observations. This results in a hypothesis. After observing new data this hypothesis can be confirmed or rejected and replaced by a new hypothesis. In observing the overt behavior of other people, hypotheses can be formulated on the internal state of mind of those people. As counterpart of human sense organs the SAM-model uses a camera and microphone. Physiology is however not a direct representation of one of the human sense organs. Some physiological signals, like heart rate and skin temperature, are observable by human sense organs, but only in a rude approximation and not detailed enough to derive behavior. The SAM-system will record physiological activity by a set of dedicated sensors, like EKG-sensors and blood pressure cuffs. The inputs of the system are data from face analysis, voice analysis, body movements and physiological signals. We will discuss them next in more details.

2.1 Stress

The concept of stress has its origins in the natural sciences. Stress was used to indicate mechanical force acting on a body, its strain being the reaction on stress. But it is now widely used in physiology and psychology. Stress is not easy to define as it involves the internal and external stimulation of an organism, which upsets its internal balance. From the viewpoint of system theory, stress is the state of threatened balance, harmony of homeostasis. It involves a series of physiological and behavioral changes that prepare the organism for the appropriate coping response. These changes are characterized by some degree of activation (arousal or emotional tension in the organism), which ensures the body to be ready for action. The physiological stress response of an organism is dominated by sympathetic arousal characterized by changes in heart rate, blood pressure, respiration rate, and muscle tension. Of these, respiration and muscle tension are likely to have effect on speech production. The physiological stress process results in overt behavior are characteristic facial expressions and typical body movements. A common definition of stress in Industrial and Organizational Psychology is the following: "A potential for stress exists when an environmental situation is perceived as presenting a demand which threatens to exceed the person's capabilities and resources for meeting it, under conditions where he expects a substantial differential in the rewards and costs for meeting the demand versus not meeting the demand. Stress, in addition to being itself and the result of itself is also the cause of itself."

Stress has a few different aspects:

- Stressors are the stressful situations, circumstances and stimuli that cause stress.
- The subjective emotional evaluation of the stressors in physical and mental respect. The outcome of this evaluation process can be positive or negative.
- At this point emotions come into play.
- Coping processes are the strategies of the individual to handle the stressful situation.
- Strains are reactions to stress. Four different reactions levels can be differentiated:
  - Cognitive level, with restricted level of perception, blocking of logical thinking and lack of concentration.
  - Emotional level, with the feelings of fear, despair, frustration and depression for example.
  - Physical level, with increase in blood pressure, heart rate, headache, stomachache.
  - A complicated stress reaction, in which mental and physical processes are involved, such as body movements and nonverbal expressions.

Research has shown that there is a strong relation between our levels of performance and attentiveness and the amount of stress we are exposed to at any time (Yerkes-Dawson curve). A certain amount of stress is necessary to perform optimally. In this phase stress is a challenge, but positively evaluated, because people have the idea that they master it. If the amount of stress however surpasses a certain threshold, stress is a dread with harmful consequences, which is evaluated as negative and can have complete dysfunctional behavior as a consequence.

Table 1. Average F0 and Jitter the 5 test samples of 6 test persons (probants)

ID	$F0_1$	$F0_2$	$F0_3$	$F0_4$	$F0_5$	J_1	J_2	J_3	J_4	J_5
1	103	103	106	108	115	.77	1.93	3.52	.00	.00
2	121	126	123	123	144	.49	1.55	1.13	.73	.50
3	111	109	111	103	-1	1.20	.82	.68	1.78	-1.0
4	98	105	113	102	99	.93	1.00	.54	.62	.58
5	116	103	125	105	113	1.53	.87	1.07	.62	.83
6	131	159	100	159	138	1.34	.85	.42	.66	1.57

2.2 Physiology

In psycho-physiological research, physiological signals are used to assess nervous activity of probants during a set of behavioral tasks. The human nervous system, especially the autonomous nervous system, appears to be differently activated during different tasks. As a result of psycho-physiological research, distinct groups of possible types of behavior can be distinguished based on nervous activity. Nervous activity can be assessed by examining physiological signals like heart rate and blood pressure [1]. From psycho-physiological theory it is also known that stressful behavior is related to special nervous activity. For example, arousal is generally associated with stress and shows an increase in blood pressure and acceleration of heart rate. The human nervous system can be divided in several parts (see Figure 1). In behavioral and stress research, the autonomous nervous system is the most important one. Stress and increased workload result in activation of the beta-adrenergic nervous system. Beta-adregenic activation is however ambiguous as it can also indicate physical labor, like using the pedals/throttle in a car. Currently the following physiological signals have used in the SAM system:

- Electrocardiogram of heart rate
- Photoplethysmograph for blood pressure and blood volume
- Chest strains for respiration rate and volume
- Galvanic skin response for sweat activity

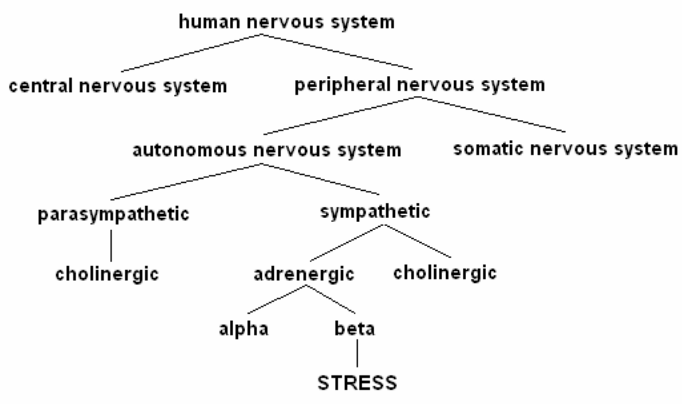

Figure 1. Description of the human nervous system

2.3 Sound

Voice analysis is based on the way of speaking and on the verbal content of messages. We discuss only the first aspect. Stress has an impact on muscles involved in speech production and respiration. Several stress features can be extracted from voice signals. A detailed description of important parameters in stress research is given in [7]:

- F0 mean: fundamental frequency of the voice signal
- F00 perturbation: slight variations in glottal cycles
- Intensity mean: energy values for a voice signal averaged over an utterance
- Speech rate: length of an utterance in time
- Zero crossings: number of times a voice signals crosses the zero line

Based on the recorded values of these parameters, stress levels can be assessed.

2.4 Facial expressions

Facial expressions are extensively analyzed by Ekman [2]. He developed a description of the human face that is based on physical and anatomical characteristics. With 44 so-called Action Units (AU) of their Facial Action Coding System (FACS) they established a parameterization of the human face that is widely used. An action unit is a facial action that cannot be decomposed into smaller facial actions. Different muscles can be responsible for the same AU, which implies that the description takes place on a level higher than the anatomical level. With a weight associated to each AU an indicator for the strength of the AU is invoked by the expression found. Stressed people show characteristic facial expressions. A desperate, wild look in the eyes, stiffed lips, blend of anger and fear are widely recognized as stress features. On the opposite the on-set of micro-sleeps can be observed by half open mouth and a drop down of the upper eyelid. Beside facial expressions other facial indicators of stress have been used like the color of the face and the appearance of sweat.

2.5 Body movements

Involuntary body movements are also indicative for the amount of stress and micro-sleeps. When the stress level is high, changes of body positions occur more often in time. By the onset of micro-sleeps the head is moving forward/ falling down. The motion detection of the SAM model is focused on processing algorithms for lower level image problems like contour detection and detection of changes in consecutive images.

3 Analysis tool

Description of the model is given in Figure 2. The overall structure shows different multimedia data streams in horizontal direction and a hierarchical layered structure in the vertical direction. The four parallel data streams are processing paths for physiological data, sound and video. Based on the information of every data stream, hypotheses are formulated. Processing of the data streams is implemented in several prototypes, which implies that every stream can operate as an independent unit. In the absence of one or more data streams, a system based on the SAM-model will still operate and result in an output.

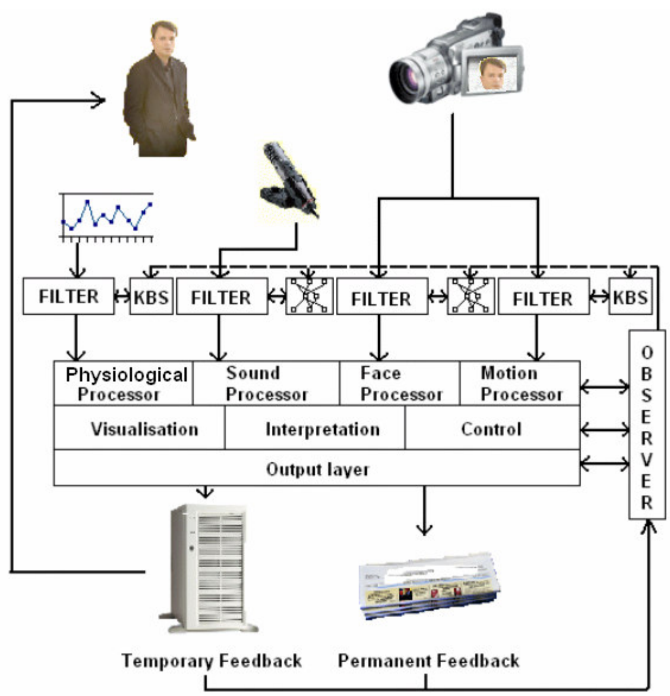

Figure 2. Description of the model

The main reason to use four independent data streams is to reduce the ambiguity in the data and to complete the data in one stream with the data of the other streams. The hypotheses are then compared and evaluated resulting in a final assumption. Finally several types of feedback can be deducted from this assumption. The main output of the system is an overall stress measure. The amount of stress is measured on an ordinal scale as the absolute amount of stress is difficult to assess. Stress has a history during which the amount of stress keeps stable, increases or decreases. So the model can only be used for stress observations in relation to time. The model consists of five layers: filtering and reduction, preprocessor, application and output layer. In the first three layers processing of the data streams takes place in parallel. A central part in the SAM is reserved for the observer. This part can interact with all the layers and is able to control every session by setting parameters in these layers.

Filtering and reduction layer

The first layer for every data stream is the filtering and reduction layer. Input of this layer consists of raw measurements. Several transformations are applied to these measurements:

- Filtering, to eliminate noise from measurements
- Outlier correction, to eliminate temporary errors in recording, data reduction, to allow for faster processing and smaller data storage capacity.

Preprocessor layer

The second layer consists of two blocks, namely the preprocessor module and the data storage module. The second layer includes functions for smoothing incoming signals and averaging over a period of time. Data storage is important for reproducibility of measurements and comparison of different implemented versions. Objective analysis can only be done in a direct comparison of results on identical input. In a real time application the data storage can provide the system with information for trend analysis to predict future behavior. The learning and adaptive character of the whole model strongly depends on storage of relevant signals.

Processor layer

This layer processes the data to derive relevant hypotheses from the signals. The processing is medium specific, because the derived hypotheses cluster in a natural way. In this layer the holistic point of view is stressed. This layer includes a database with stress prototypes, stress templates and other stress patterns. Pattern recognition techniques are used to match the input features with stored patterns. On this level a medium specific stress measure is computed.

Application layer

The application layer is structured as a rule-based system. The derived hypotheses in the processor layer are tested for consistency. A stressful emotion or an involuntary movement for example is assumed to have specific physiological reaction patterns. This layer has a database with specific knowledge about the relations between patterns for different media. Finally an interpretation and feedback are generated.

Output layer

This layer produces actual feedback. The types of feedback are derived for all active applications: temporary and permanent feedback. Temporary feedback is given to the observant to show first results and give opportunity to adapt to a situation. The permanent feedback is the basis for monitoring during successive sessions and an input for the learning process of he system. It can be used to check progress during successive training sessions.

Observer

Final and central parting the SAM-system is the observer. The observer represents the trainer or observer during training and test sessions. Interaction is possible between the observer and all parts of the system. By this interaction the observer is able to control the course of the session.

4 Experimental setting

The experiments were carried out in a lab environment. The probants were 90 male students from the Department Technical Informatics. The students had to play a well-known race car computer game. At every turn they have to speak left/right depending of their direction. The probants had no access to the acceleration of the car and the brakes function not all the time. The speed was increased during the game session and was supposed to induce increasing stress.

5 Results

5.1 Physiological prototype

In the filtering and reduction layer of the SAM system noise filtering is applied to incoming measurements, followed by outlier correction and feature abstraction. Finally features are related to some basal value, because every person has a different heart rate, blood pressure; these basal values are recorded during some rest or relaxation period. Besides storing features for later use, the pre-processor layer diminishes temporary and cyclic perturbations as they can be found in physiological data. For instance heart rate is lower during inhalation and higher during exhalation. The intermediate representation used by the processor layer to derive hypothesis about stress levels is human nervous activity, as was mentioned above. Several applications are built into the application layer. First features are projected on a two-dimensional plane. Heart rate is mapped on one axis and blood pressure on the other axis. In this way regions can be labeled by nervous activity: parasympathic, and alpha-and beta-adrenergic sympathetic enhancement or inhibition. Only cholinergic sympathetic activity cannot be monitored by this projection. Another application is the behavioral compass. This indicates the direction in which behavior will be changing in the near future. Finally a stress-meter is integrated in the prototype. Several areas in the one-dimensional space correspond to certain stress levels and these are visualized as a kind of "econometer" (see Figure 3). Important signals recorded during our research were systolic blood pressure (SBP), diastolic blood pressure (SBP), inter beat interval (IBI) and galvanic skin level (GSL). To apply univariate statistics, signals are averaged over every condition. The means and standard deviation of all variables over the first 6 conditions are displayed in the next tables 2 and 3.

Table 2. The means of the four variables on 6 conditions

Conditions	IBI	SBP	DPB	GSL
C1	909.36	134.04	74.10	12.28
C2	930.60	131.02	72.45	12.48
C3	921.13	133.39	73.20	12.52
C4	925.52	132.92	73.79	12.52
C5	880.22	146.37	80.13	19.21
C6	893.07	146.76	80.93	18.99

Table 3. The Standard Deviations of the 4 variables on 6 conditions

Standard deviation	IBI	SBP	DPB	GSL
C1	129.91	22.39	13.40	8.33
C2	132.04	21.46	13.12	8.57
C3	132.50	22.12	13.42	9.22
C4	136.14	23.17	14.54	9.52
C5	127.15	24.96	16.11	11.43
C6	133.52	25.30	16.85	11.98

5.2 Sound prototype

Noise filtering is very important as background noise can completely disguise the original voice signal. Noise filtering is the most important task of the filtering and reduction layer. The other important task is feature extraction. At this moment we are able to extract the following parameters: F0, F00 (jitter), frequency spectrum, spectrogram, energy, magnitude and zero crossing. The preprocessor layer takes care of storing samples and average features over speech utterances. The processor layer has to compute the relation between stress and voice features. It is still under construction. The application layer consists of two applications. The first covers the visualization of speech signals. The second application is a voice stress-o-meter. This stress-o-meter is based on a linear combination of fundamental frequency and jitter.

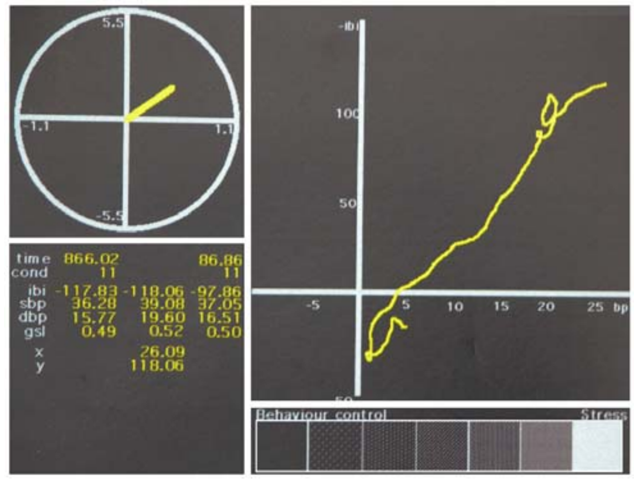

Figure 3. Stress track, compass and stress-o-meter

From table 4 it can be concluded that the fundamental frequency and jitter correlated with the stress level as induced by increasing speed. This is a result of the increasing tension in the vocal folds, causing less muscle vibration and therefore less variation in fundamental periods.

Table 4. Averages of the number of cycles measured for F0 and jitter averages

Speed Condition	Mean	St.dev	Min	Max	N
Low	8.64	5.12	1	22	87
Below average	9.45	6.53	1	34	83
Average	10.25	6.27	1	41	76
Above	10.89	5.92	1	34	84
High	12.34	7.84	1	54	77

References

[1] Berntson, G.G., Cacioppo J.T. and Quigley K.S. "Autonomic Determinism: The Modes of Autonomic Control, The Doctrine of Autonomic Space, and the law of Autonomic Constraint." Psychological review 98, no. 4: 59-487 1991

[2] Ekman, P. and W. V. Friesen "Unmasking the face. A guideline to recognizing emotions from clues" Prentice Hall, New York 1975

[3] Pantic M. and Rothkrantz L.J.M. "Facial action recognition for facial expression, Analysis from Static face images", in IEEE Transaction on Systems, Man and Cybernetics, vol. 34, no. 3, June 2004

[4] Pantic M. and Rothkrantz L.J.M. "Towards an affect-sensitive multimodal human-computer interaction" in Proceedings of the IEEE vol. 91, pp. 1370-1390 Sept. 2003

[5] Pantic M. and Rothkrantz L.J.M. "Expert system for automatic analysis of facial expressions" Image Vision Computing Journal vol 18, no. 11, pp. 881-905, 2000

[6] Pantic M. and Rothkrantz L.J.M. "Automatic analysis of facial expressions: The state of the art" IEEE Trans. Pattern Analysis, Machine Intell vol 22, pp. 1424-1445, Dec. 2000

[7] Scherer, K.R. In: 'Emotions: Theory, research and experience'. Volume 4: "The measurements of emotions" Plutchik, R. and H. Kellerman, Academic Pres, San Diego, USA, 233-259 1989