Exploring Smart Agents for the Interaction with Multimodal Mediated Environments

Multimodal Technologies and Interaction

MDPI

Article

Exploring Smart Agents for the Interaction with Multimodal Mediated Environments

Robert Richer [1,*], Nan Zhao [2], Bjoern M. Eskofier [1] and Joseph A. Paradiso [2]

[1] Machine Learning and Data Analytics Lab., Department of Computer Science, Friedrich-Alexander-Universtität Erlangen-Nürnberg (FAU), Carl-Thiersch-Str. 2b, 91052 Erlangen, Germany; bjoern.eskofier@fau.de

[2] Responsive Environments Group, MIT Media Lab., 75 Amherst St, Cambridge, MA 02139, USA; nanzhao@media.mit.edu (N.Z.); joep@media.mit.edu (J.A.P.)

* Correspondence: robert.richer@fau.de; Tel.: +49-(0)9131-85-20162

Received: 30 April 2020; Accepted: 3 June 2020; Published: 6 June 2020

Abstract: After conversational agents have been made available to the broader public, we speculate that applying them as a mediator for adaptive environments reduces control complexity and increases user experience by providing a more natural interaction. We implemented and tested four agents, each of them differing in their system intelligence and input modality, as personal assistants for *Mediated Atmospheres*, an adaptive smart office prototype. They were evaluated in a user study ($N = 33$) to collect subjective and objective measures. Results showed that a smartphone application was the most favorable system, followed by conversational text and voice agents that were perceived as being more engaging and intelligent than a non-conversational voice agent. Significant differences were observed between native and non-native speakers in both subjective and objective measures. Our findings reveal the potential of conversational agents for the interaction with adaptive environments to reduce work and information overload.

Keywords: ubiquitous computing; smart office; conversational agents; smart agents; chatbot; human–computer interaction; adaptive environments; multimodal interaction; mediated atmospheres

1. Introduction

Smartphones have become an essential part of our daily life and are deeply embedded in our everyday actions, in a personal as well as in a professional environment [1]. We use our smartphones about 85 times a day, spending in total more than 5 h with the device [2]. Interacting with a smartphone requires us to use our hands and easily distracts us from our other activities [1], therefore companies like Apple (*Siri*, 2011), Google (*Google Now*, 2012), Microsoft (*Cortana*, 2015), and Amazon (*Alexa*, 2015) have released smart voice assistants that are integrated into smartphones, computers, or IoT devices and enable hands-free and more unobtrusive interaction. In 2016, Google announced *Google Assistant* as the successor of *Google Now*. It was designed as an intelligent personal assistant that engages the user in a two-way conversation, but also offers multimodal input via voice and text [3]. Research groups have been developing agents that answer questions conversationally for the past two decades [4]. However, those companies have made them available for the broader public, following the vision of the HCI community that the importance of voice agents will considerably increase as they will become more conversational [5].

We believe that the emerging potential of such conversational agents offers new possibilities for smart environments by facilitating interaction and to make them more adaptive—a necessity for environments in order to appear intelligent, according to Mozer et al. [6]. By equipping the environment with pervasive and ubiquitous technologies, it obtains information about itself, and the

actions and preferences of the users. Through sensing their requests and responding accordingly, users may perform those complex, seemingly intelligent tasks automatically [7] and create a symbiosis between the space and the occupant for providing meaningful interactions with our surroundings.

As one scenario of an intelligent, adaptive environment, we use *Mediated Atmospheres*, a modular workspace proposed by Zhao et al. [8]. It aims to dynamically influence the users' abilities to focus on a specific task or to recover from a stressful situation. Therefore, multimodal media—lighting, video projection, and sound—is used to transform the appearance of the workspace. It is leveraged to address concerns in contemporary, open-plan workplaces where recent reports showed a decline in the satisfaction of office workers, leading to reduced productivity and motivation. In contrast, employees reported that having more freedom in choosing their workplace leads to a higher level of satisfaction, innovation, and effectiveness [9]. Utilizing *Mediated Atmospheres*, different environments can be projected into the room according to the occupants' preferences or different work scenarios in order to foster creativity, productivity, and relaxation [8].

In order to unleash the full potential of such an adaptive environment, we do not want the user experience to be compromised by a cumbersome and tedious interaction with our workspace, e.g., by manually adjusting lighting and sound to change its appearance. We rather envision a system acting as a mediator between us and our environment, thus reducing our work and information overload [10]. It should be capable of understanding our requests and seamlessly translate them into a manifestation of lighting, video, and sound.

However, nowadays, conversational agents still suffer from a "gulf of evaluation" between users expectations and system operation [11]. Naturally, users have higher expectations towards a system that pretends to be capable of conducting a conversation, and hence tend to be more disappointed if the system provokes wrong actions [11]. Nonetheless, we speculate that the "gulf of evaluation" depends on the input modality, and is wider for voice-based input than for text-based input.

Therefore, this paper explores the application of different agents as personal assistants for the *Mediated Atmospheres* framework as an example for a multimodal mediated environment. We believe that each interface has its strengths and weaknesses, and none of them will offer a perfect solution for the interaction with a smart office prototype. We speculate that each interface is applicable for different scenarios in the daily work flow in a smart office, depending on the current task and the user preference. To evaluate the interplay between the agents and the multimodal mediated environment we present and compare four different systems that are integrated into our smart office prototype, intended to increase both system usability and user experience by providing a natural way of interaction. By creating and evaluating agents that differ in their level of system intelligence and input modality, we want to assess if these agents are superior to conventional interaction modalities, like a smartphone-based graphical user interface. Furthermore, we are interested to find out whether we can observe differences between native and non-native speakers. Especially for natural language interfaces, we speculate that these differences are more pronounced compared to other interfaces, because possible false recognition might influence user experience or perception. Finally, by synthesizing our results, we want to give recommendations for what an effective agent might look like for the context of smart, adaptive environments.

2. Related Work

In the past years, researchers have equipped the interior of buildings with pervasive and ubiquitous technologies in order to control specific properties like lighting [12–14], glazing regulation [15], HVAC control [16], sound [17], or information display [18–20].

Other research groups have created whole adaptive ambient environments to enable life-size telepresence [21], or to transform its appearance and physical shape according to physiological data of the occupant [22]. Another example is the ambientROOM that utilizes subtle cues of sound, light, or motion to provide meaningful information to the user [23]. Similarly, context-aware systems have also been explored within the last two decades. For instance, researchers have created the Aware Home, an entire home equipped with ubiquitous computing technologies in order to create a living

laboratory that is as realistic as possible [7]. Other examples that aimed to increase convenience are the Neural Network House [24], or the Adaptive House [25].

In order to evolve from adaptive, augmented rooms to intelligent environments [26], agents have been integrated into the built environment. They are supposed to learn the occupants' preferences and habits and, then, execute appropriate actions automatically. For instance, MavHome is an agent-based smart home with algorithms to predict the inhabitants' actions [27]. Other smart home applications see the benefit of agents in power management [28], or to mediate the interaction with a smart grid [29]. In smart offices, agents have been integrated for a virtual secretary [30], or to accompany and lead persons in an office building [31]. With increasing popularity of smart agents and environments, people have analyzed them in terms of acceptance and trust [32], dominance and cooperativity [33], or personality traits [34].

The idea of an agent, an artificial intelligence designed to maintain a conversation with a human, has already been envisioned by Alan Turing in the 1950s [35] and was successfully implemented for the first time by Joseph Weizenbaum with ELIZA [36]. It is considered the first "chatbot", a text-based conversational agent simulating a natural human interaction between the computer and a user. Recently, conversational agents have been created for the touristic [37] and customer service domain [38], or in social networks like Facebook or Twitter [39]. Additionally, agents have found wide-spread applications in healthcare applications and clinical psychology, as indicated by recent reviews from Laranjo et al. [40], Montenegro et al. [41], or Provoost et al. [42]. The findings of the reviews highlight the potential of conversational agents in specialized areas, especially in the health domain. They not only increase productivity by assisting medical professionals during consultation and documentation, but also increase well-being by interacting and engaging with patients during their treatment. Through rapid technological advancements—especially the ability of understanding unconstrained natural language input—it has even become hard to distinguish between humans and chatbots in social networks [43,44]. For that reason, we believe that smart conversational agents have also the potential to be applied in multimodal environments in a similar way to reduce work and information overload during everyday work activities in an office scenario.

Researchers like Schaffer et al. [45], Weiss et al. [46], or Silva et al. [47] presented multimodal input solutions for applications like smart homes or smartphones using gestures, graphical user interfaces, or voice. Additionally, Kühnel et al. [48] analyzed the applicability of established questionnaires for evaluating multimodal input systems and concluded recommendations of which methods to use in order to compare different systems which also influenced the choice of our evaluation measures. However, none of them investigated different levels of system intelligence. Furthermore, they did not consider possible differences between native and non-native speaker. This was investigated by Pyae and Scifleet in a study analyzing the usability of voice user interfaces to better understand usage differences between native and non-native speaker. Their findings reveal that native speakers experienced better and more positive user experiences and also show that English language proficiency is an important factor for the interaction with voice user interfaces [49].

3. System Description

We designed four systems that differ in the input modality (graphical vs. voice vs. text) and the system intelligence (basic vs. advanced), as visualized in Figure 1. As shown in a top-level description of our setup in Figure 2, the agents communicate with *Mediated Atmospheres*, a smart office prototype that was introduced in previous work [8].

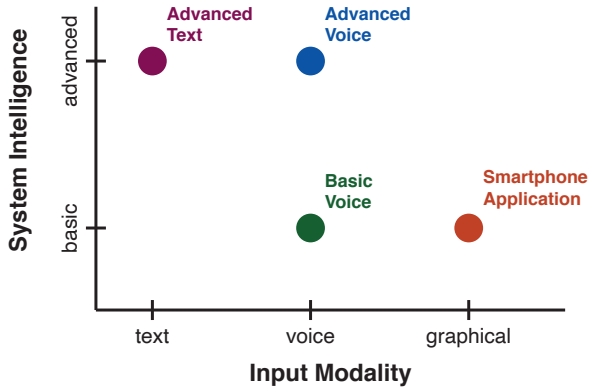


Figure 1. Overview of systems for the interaction with the smart office prototype.

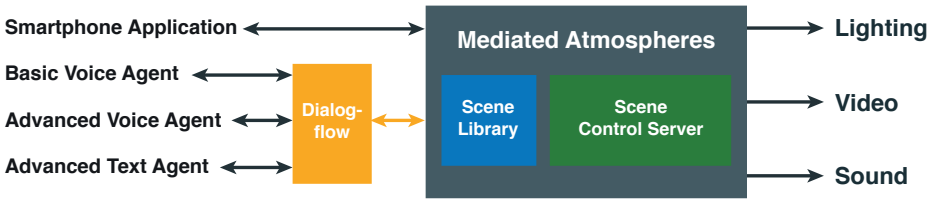


Figure 2. Top-level description of agent integration into *Mediated Atmospheres*.

3.1. Mediated Atmospheres

Mediated Atmospheres is a modular workspace capable of digitally transforming its appearance using controllable lighting, video projection, and sound. We combine all digital output capabilities to design a library of scenes, multimodal virtual environments, with each of them possibly having a different influence on our physiology or behavior [8]. Examples of the workspace, transformed into different scenes, are shown in Figure 3.

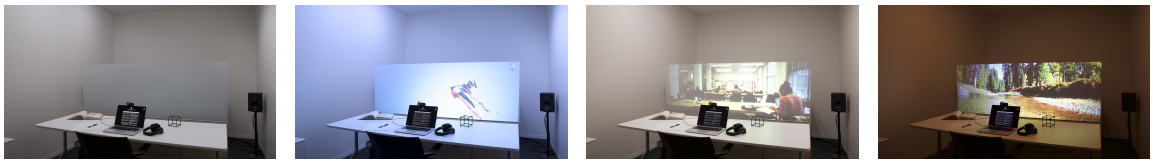


Figure 3. *Mediated Atmospheres* office space with different scenes.

The scene library currently features more than 30 elements, ranging from nature landscapes, like forests, beaches, mountains, over urban scenes, like walking through a city or train rides, to indoor spaces, like libraries or cafés. The scenes were selected to cover a wide range of environments. Each of them are identified by a descriptive name, and contain video and sound data, as well as lighting configurations. Furthermore, we defined different scene properties, that will later be used by the agents to filter scenes, like:

- *Color Temperature*: Color temperature of the scene's lighting configuration in Kelvin
- *Brightness*: Brightness of the scene's lighting configuration, measured as the horizontal illumination on the desk in lux
- *Primary Color*: Dominant color of the scene's video content
- *Keywords*: Keywords describing the scene's video content

The scene library was implemented as an Amazon dynamoDB database, a scalable and non-relational database web service, that offers APIs for platforms like Python, Node.js, or Android [50].

For controlling the office space in real-time, the *Mediated Atmospheres* system features a *Scene Control Server* with a Websocket API to dynamically control the appearance of the workspace from different applications (as visualized in Figure 2).

3.2. Smartphone Application

The smartphone application was developed for Android-based mobile devices using the Android SDK 9.0 (API level 28). It is able to communicate with the *Scene Control Server* to change the current scene and is connected to the *Scene Library* for fetching scene information. The user interface is based on the "What You See Is What You Get" (WYSIWYG) principle, allowing users to see preview images for each scene, filter scenes by their properties using sliders (see Figure 4), and change the environment by selecting a scene. An overview of features available in the smartphone app is depicted in Table 1.

Table 1. Feature range of agents. x = available, – = not available.

	Agent			
Feature/Action	**Smart-Phone**	**Basic Voice**	**Advanced Voice**	**Advanced Text**
Turn on/off	x	x	x	x
List scenes	x	x	x	x
Current scene	x	x	x	x
Switch scene based on:				
- Descriptive name	–	x	x	x
- Scene properties (color temperature, brightness, etc.)	x	–	x	x
- Comparison with current scene (warmer, brighter, etc.)	–	–	x	x
- Keywords describing scene content (forest, beach, etc.)	–	–	x	x
Context-awareness (information about occupant and previous interactions)	–	–	x	x
Recommend scene based on:				
- Current weather information	–	–	x	x
- Current time of day	–	–	x	x
- Desired mental state of occupant (focus/relaxed)	–	–	x	x

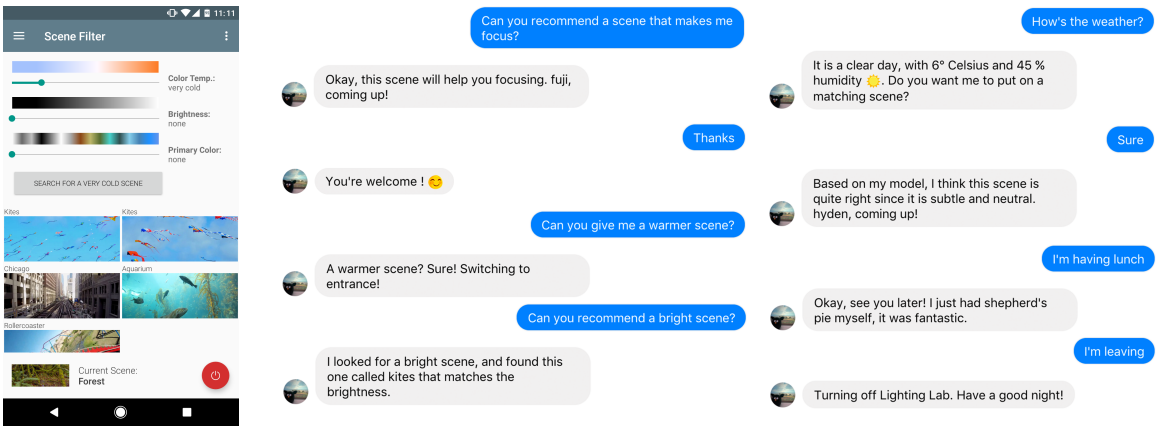


Figure 4. Examples of user interface for interaction with *Mediated Atmospheres*. **Left**: Graphical User Interface of *Smartphone Application*; **Right**: Example dialog between user and *Advanced Text Agent*.

3.3. Smart Agents

All agents were implemented in Dialogflow, a platform for building conversational agents for different applications (https://dialogflow.com). It uses *Intents* for representing the mapping between the *User Input* and the *Action* the system should perform, and provides a webhook integration. For this work, we used a Node.js backend to perform further logic, create more diversified agent responses, access the scene library, and connect the agents to external APIs. The agent designed in Dialogflow can be integrated into different platforms. For the voice agents, we used the *Actions on Google* platform, which deploys the agent on Google Home (Google Inc., Mountain View, CA, USA), a wireless smart speaker. It runs the Google Assistant, an intelligent personal assistant designed to engage the user in a two-way conversation (https://blog.google/products/assistant/io-building-next-evolution-of-google). Once deployed on Google Home, an agent can be invoked by saying "Okay Google, let me talk to {*agent_name*}".

For the text agent, we used the Facebook Messenger platform to build a chatbot for a Facebook page, allowing users to directly text the agent with the desired action, without any invocation necessary (https://developers.facebook.com/docs/messenger-platform/getting-started/quick-start). In order to make the agent more engaging, we included emojis into its text responses (see Figure 4 for an example dialog).

3.3.1. Basic Voice Agent

This voice agent was designed to cover the basic requirements of an agent for *Mediated Atmospheres*, so that we end up with the features listed in Table 1. The agent is non-conversational, meaning that the interaction with the user is based on a simple Request–Action scheme with static text responses, rather than engaging the user in a dialog. As it has no information about scene properties, users have to directly tell the agent which scene to switch to by using its descriptive name, either by having prior knowledge about certain scenes, or by asking the agent to list a subset of scenes.

3.3.2. Advanced Voice and Text Agent

The advanced agents exceed the basic agent in terms of both feature range (see Table 1) and context-awareness. As shown in an example dialog in Figure 4, the advanced agents were implemented as conversational agents, incarnating a personal assistant for the smart office. Therefore, they talk about themselves in first person and ask occupants for their names in order to personally address them and to remember previous interactions. Furthermore, we designed the advanced agents to offer a wider range of possible text responses for more diversified conversations. Text responses consist of randomly selected text modules from a database, which are concatenated and modified depending on the current context, such as the current time or whether the user is leaving the room at the end of his work day or just to have a break. For all possible actions, we designed interaction dialogs in Dialogflow to cover as many contingencies as possible. Figure 5 shows the possible interaction flow for the *Current scene* action.

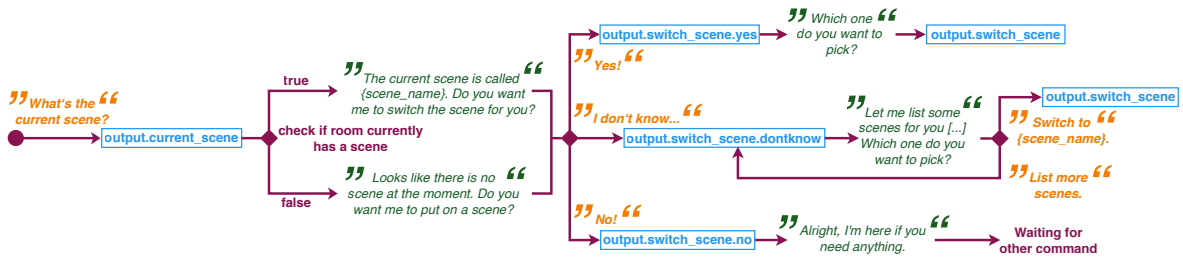


Figure 5. Possible interaction dialog with advanced agent for *Current scene* action. Orange: User input; Green: Agent output; Blue: recognized intent.

Because both advanced agents have information about scene properties, users can not only switch to a scene by specifying the descriptive name, but are also able to select scenes by more abstract queries,

like "Can you take me to a nature scene?", or "Can you put on a warm and bright scene?". It is further possible to switch to a scene with properties relative to the current scene, like a scene with warmer or colder color temperature. However, these features still require the users to map their scene expectation to the different properties. Therefore, we wanted the agent to recommend scenes based on the current context, or the users' preferences:

- *Weather*: Recommending a scene according to the current weather was based on mapping real-time weather information (obtained from the Dark Sky API (https://darksky.net/dev)) to scene properties. We mapped the general weather situation to the scenes' color temperature to approximate the current lighting situation outside. Additionally, because high levels of humidity were found to be a key factor for concentration loss and sleepiness [51], we counteracted this phenomenon by recommending scenes with higher brightness.
- *Time*: Color temperature of natural light changes over the day. It is colder in the morning and at noon, and warmer during afternoon and evening hours [52]. Therefore, scenes with different color temperatures are recommended for different day segments to match the human circadian rhythm [53].
- *Desired mental state*: The advanced agents are aware of current occupant in the workspace, allowing them to recommend scenes to users based on their preferences, like scenes that helped them being more focused or more relaxed. To set these preferences users can create their own models using a website. The agent was then able to access these personal scene models.

4. Evaluation

4.1. Experiment Design

Our population of interest is mainly knowledge workers. Therefore, we recruited university students, researchers, and local office workers as subjects for our experiment. $N = 33$ people (61% Female) participated in the experiment, with an average age of 27.5 ± 3.6 (M $\pm$ SD). 52% of the participants named English as their native language. The experiment consisted of two parts: *Agent Exploration* and *Task Completion*.

4.1.1. Agent Exploration

The first part of the experiment was conducted as a within-subject design. For this, each participant interacted with every agent: Smartphone Application (*SA*), Basic Voice Agent (*BV*), Advanced Voice Agent (*AV*), and Advanced Text Agent (*AT*). For each agent, they were asked to complete the following list of tasks using all features available for the respective agent:

1. *Turn Mediated Atmospheres off because you're about to get lunch*, followed by *Turn Mediated Atmospheres on after returning from lunch*
2. *Find a scene that has warm color temperature*
3. *Find an indoors scene*
4. *Find a scene that helps you focus*
5. *Find a scene that matches the current weather*

Both the order of agents and the order of tasks for every agent were randomized to avoid possible adaptation effects. The *Agent Exploration* part was mainly intended to make the subjects familiar with the usage of each agent and allow them to explore the different ranges of features. We further collected subjective responses in a survey to measure overall usability, perception of intelligence, and engagement, as well as perception of trust and control. These measures are explained in detail in Section 4.3.

4.1.2. Task Completion

After the first part, all subjects should have the same knowledge on how to communicate with the agents, which allows us to measure quantitative information about the agent interaction in the second part of the experiment. For that, we chose a between-subjects design where each participant was randomly assigned to one agent to allow a better assessment of objective measures without possible habituation effects. Additionally, we wanted to limit the duration of this explorative study to a reasonable length. In total, nine subjects were assigned to the *Advanced Voice Agent*, and eight subjects were assigned to each of the other agents, respectively.

For the experiment, participants were asked to complete the following list of tasks utilizing the agent's features:

1. *Find a city scene*
2. *Find a scene that shows mountains*
3. *Find a bright and blue scene*
4. *Find a scene that is warm and shows a forest*
5. *Find a scene that matches the current time of day*
6. *Find a scene that helps you relax after a rough day*

The tasks were selected to represent the different features from Table 1 equally. For each task, we collected objective measures like task completion time, number of interactions, and agent recognition rate. These measures are explained in detail in Section 4.4.

4.2. Procedure

Both parts of the experiment were performed in the *Mediated Atmospheres* office space and consecutively in one sitting. Each sitting began with a tutorial, during which the study personnel described the concept of *Mediated Atmospheres* and explained the study protocol. It was followed by an introduction into the different agents, their range of features, and the way of interacting with them. Subsequently, participants were left alone in the workspace, and a website showing the task list for the currently selected agent guided them through the first part of the experiment. After completing all tasks with one agents, subjects were asked to provide feedback in form of a questionnaire and advance with the next agent. This was repeated for every agent.

For the *Task Completion* part, subjects were left alone in the workspace, again with a website showing a list of tasks to complete with the agent assigned to them. Since we were interested in collecting objective measures for this part, participants were asked to press a button once they successfully accomplished a task to advance with the next one.

4.3. Subjective Measures

4.3.1. Overall Usability

For a subjective measure of the usability for each agent, we used the System Usability Scale (SUS) [54]. It consists of a 10 item questionnaire with a five-level Likert scale (1 = Strongly Disagree, 5 = Strongly Agree), and yields a high-level subjective view of the usability as a linear scale of 0–100. We chose the SUS scale over questionnaires particularly designed for speech interfaces, such as SASSI [55], because it is independent from the input modality and allows to compare the different systems [48]. At the end of Part I of the experiment, we further asked the participants to directly compare the agents by ranking them according to their personal preference, with 1 being the most favorite, and 4 being the least favorite agent.

4.3.2. Intelligence and Engagement

We were interested in how intelligent and engaging the agents were found by the users. Therefore, we asked the study participants to rank intelligence and engagement for all the agents on a five-level

Likert scale (1 = Not Intelligent/Engaging, 5 = Very Intelligent/Engaging). Additionally, we asked the participants to write down the reasons why the agents were found to be (not) intelligent or engaging, respectively.

4.3.3. Trust and Control

As some of the agents recommended scenes, and, hence, autonomously made decisions, the perception of trust in the system and the feeling of being in control can be affected by the rate of automation error or misautomation. Therefore, we were interested in measuring the perception of trust and control towards the different agents. We adopted a survey proposed by Hossain et al. [32] and modified it to a five-level Likert scale to fit to the rest of our questionnaire. It consisted of an equal number of questions having positive and negative trust and control implications.

4.4. Objective Measures

4.4.1. Task Completion Time

Task completion time was computed as the absolute time participants needed to find a scene that was in their opinion appropriate for the given task.

4.4.2. Number of Interactions

For the *Smartphone Application*, the number of interactions was defined as the number of times users pressed the *Search for scenes* button per task. For the voice and text agents, it was defined as the number of times users invoked one of the agents' actions from Table 1.

4.4.3. Recognition Rate

For the *Smartphone Application*, recognition rate was assumed to be 100%, because users directly changed scenes without having an agent in between. For the voice and text agents, recognition rate per task was defined as the number of correctly recognized and mapped actions divided by the total number of interactions needed to accomplish the task.

4.5. Statistics

For statistical analyses, homogeneity of variances were assessed by the Levene test. Greenhouse-Geisser corrections were applied if the assumption of sphericity (indicated by Mauchly Test), was violated. We then conducted t-tests to determine group differences (native vs. non-native speaker) and repeated-measurement ANOVAs to assess differences between the agents. As post-hoc test the Tukey Honestly Significant Difference (HSD) test was used. Significance level α was set to 0.05. Effect sizes are reported as Cohen's d for t-tests and η_p^2 with 95 % confidence intervals for ANOVAs.

5. Results

5.1. Subjective Measures

5.1.1. Overall Usability

The highest SUS score was achieved by *SA* with 79.8 ± 20.0 ($M \pm SD$), followed by *AT* with 76.3 ± 20.2, the *AV* with 61.3 ± 20.6, and *BV* with 47.0 ± 24.4. Only *SA* and *AT* achieved SUS scores above 68, which is considered "above average" [54]. Native speakers rated the overall usability of *AV* considerably lower than non-native speakers (56.0 ± 20.6 vs. 66.9 ± 19.8), whereas no differences were observed for the other agents.

Similarly, *SA* was ranked as the best input modality by 39.4 % of the users, followed by *AT* (36.4 %), *AV* (15.1 %) and *BV* (9.1 %). The majority (57.6 %) of study participants ranked *BV* as the worst input modality.

Table 2 shows the response to the question whether the agents met the participants' expectations. For both native and non-native speaker, *BV* was the only input modality for which expectations were not met, i.e., where a negative scoring was achieved.

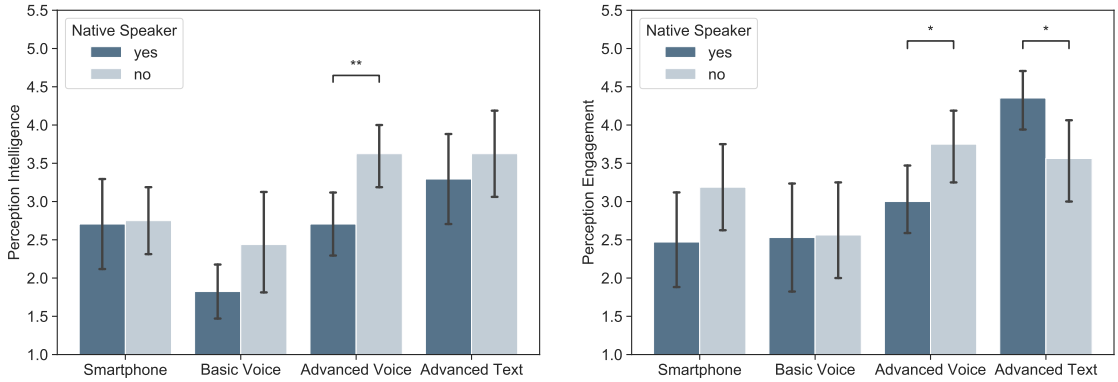
Table 2. Results of question whether expectations of subjects towards agents were met. Values are reported as mean ± standard deviation. (M ± SD). Negative values (indicating that expectations were not met) are denoted in **bold**. *SA* = Smartphone Application, *BV* = Basic Voice Agent, *AV* = Advanced Voice Agent, *AT* = Advanced Text Agent.

	SA	BV	AV	AT
Native Speaker	1.00 ± 1.37	$\mathbf{-0.18 \pm 1.38}$	0.12 ± 1.32	0.47 ± 1.50
Non-native Speaker	1.12 ± 0.89	$\mathbf{-0.62 \pm 1.41}$	0.25 ± 1.06	1.06 ± 1.34

5.1.2. Intelligence and Engagement

Repeated-measurements ANOVA revealed significant differences between the agents in the perception of intelligence and engagement ($F(3,96) = 12.31, p < 0.001, \eta_p^2 = 0.278$ and $F(3,96) = 8.766, p < 0.001, \eta_p^2 = 0.215$), respectively. Post-hoc testing showed significantly lower perceived levels of intelligence between *BV* and all other agents as well as significantly lower levels of engagement between *BV* and the two advanced agents. Additionally, *AT* showed significantly higher perceived levels of engagement than all other agents.

Figure 6 visualizes the perceived levels of intelligence and engagement for native and non-native speaker, separately. T-testing revealed significantly lower perceived levels of intelligence and engagement for *AV* among native speakers compared to non-native speakers (intelligence: $t(30.98) = -2.926$, $p = 0.006$, $d = -1.018$, engagement: $t(30.88) = -2.153$, $p = 0.039$, $d = -0.750$). In contrast, native speakers perceived *AT* to be significantly more engaging than non-native speakers ($t(27.73) = 2.220$, $p = 0.035$, $d = 0.780$).



(a) Intelligence **(b)** Engagement

Figure 6. Perceived levels of *Intelligence* and *Engagement* per agent for native and non-native speaker. Error bars represent the 95% confidence interval. *Note:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.1.3. Trust and Control

Overall, *SA* achieved the highest positive and lowest negative levels of trust and control (positive: 4.15 ± 0.92, negative: 1.42 ± 0.80), followed by *AT* (positive: 3.92 ± 0.91, negative: 1.63 ± 0.81), *AV* (positive: 3.20 ± 0.87, negative: 2.15 ± 0.88), and *BV* (positive: 2.78 ± 1.15, negative: 2.34 ± 1.18). Repeated-measurement ANOVA revealed significant differences between the agents for both positive and negative implications (positive: $F(3,96) = 17.113, p < 0.001, \eta_p^2 = 0.348$, negative: $F(3,96) = 11.123, p < 0.001, \eta_p^2 = 0.258$ after Greenhouse-Geisser correction), respectively. Post-hoc testing

showed that both *SA* and *AT* scored significantly higher levels of trust and control compared to the other two agents.

5.2. Objective Measures

5.2.1. Task Completion Time

Task completion times for Part II are visualized in Figure 7. Mean task completion times, normalized to the task completion times of *SA*, are further listed in Table 3. Results yield that tasks were on average performed fastest with *SA* and that the mean task completion times of the two voice agents are significantly higher (BV: $t(10) = -2.754$, $p = 0.020$, $d = -1.590$, AV: $t(10) = -3.223$, $p = 0.009$, $d = -1.861$)). Additionally, all agents showed significantly higher standard deviations in the task completion times compared to *SA* (BV: $t(10) = -4.078$, $p = 0.002$, $d = -2.355$, AV: $t(10) = -4.350$, $p = 0.001$, $d = -2.511$, AT: $t(10) = -2.335$, $p = 0.042$, $d = -1.348$). On average, users needed 1.26 and 1.60 times longer with *AT* and *AV*, respectively, whereas *BV* had a 2.02 times higher mean task completion time. Furthermore, results show that participants using *BV* had particular trouble finding a "Mountain" scene (353.9 %), whereas finding a "Warm Forest" scene was apparently difficult when using one of the advanced agents (*AV*: 234.3 %, *AT*: 190.0 % higher task completion times, compared to *SA*). However, on average, two out of the six tasks ("City" and "Relaxing") were performed faster with *AT* than with *SA*.

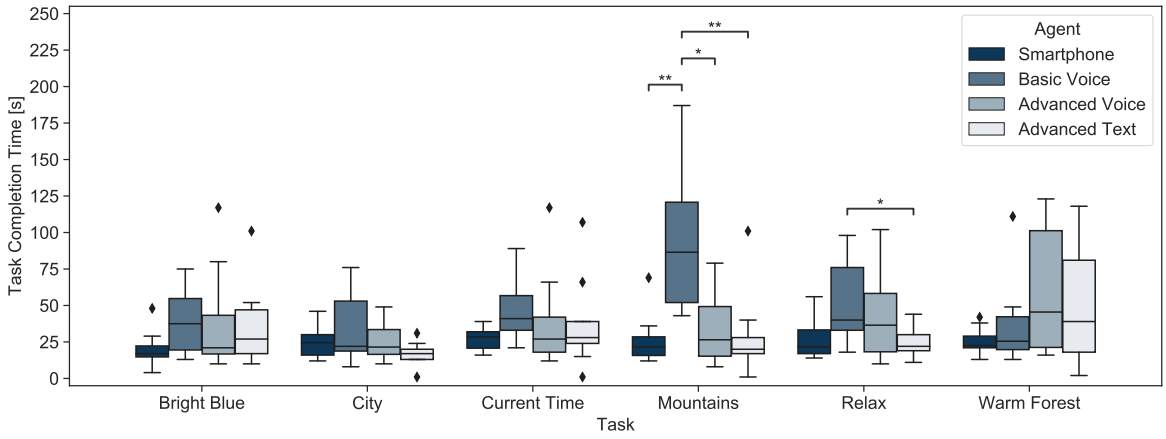


Figure 7. Task completion time per task and agent during Part II. *Note:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3. Mean task completion time and mean number of interactions per task and agent during Part II of the experiment. Values are reported in percent, normalized to the task completion time and the number of interactions of *SA*, respectively. *SA = Smartphone Application, BV = Basic Voice, AV = Advanced Voice, AT = Advanced Text.*

		Bright & Blue	City	Current Time	Mountains	Relaxing	Warm Forest
Task Completion Time (in %)	*SA*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*
	BV	190.2	143.7	171.5	353.9	193.5	147.5
	AV	189.6	105.6	142.5	129.0	164.0	234.3
	AT	177.2	67.2	134.7	106.9	92.6	190.0
Number of Interactions (in %)	*SA*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*
	BV	150.0	90.5	110.0	182.1	95.8	158.8
	AV	95.2	76.2	75.6	63.5	81.5	183.0
	AT	107.1	57.1	75.0	50.0	58.3	135.3

5.2.2. Number of Interactions

Number of interactions per task are visualized per agent in Figure 8. Average number of interactions, normalized to the number of interactions of *SA*, are further listed in Table 3. On average, both advanced agents required less interactions to complete the same tasks than *SA* (*AV*: 9 % less, *AT*: 25 %), whereas participants using *BV* required 31 % more interactions.

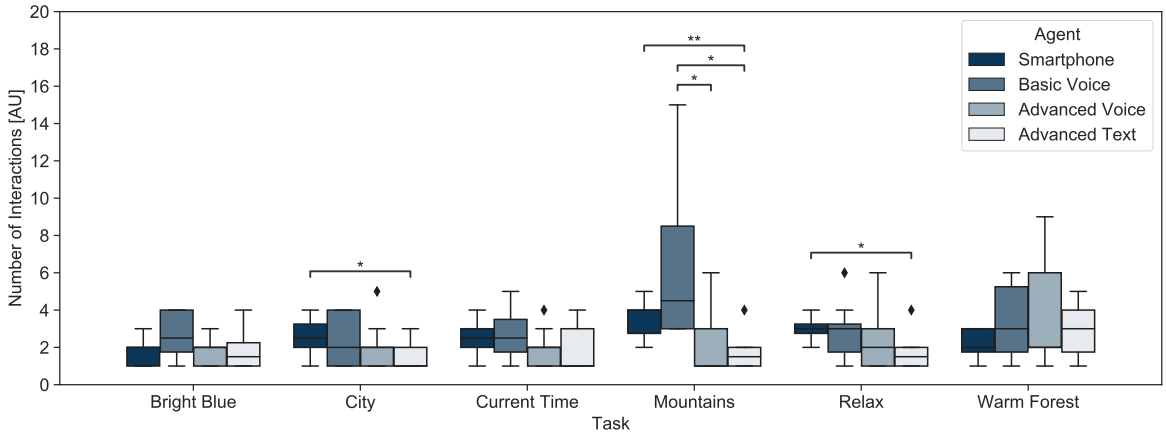


Figure 8. Number of interactions per task and agent during Part II. *Note:* * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5.2.3. Recognition Rate

Apart from *SA*, the highest recognition rate was achieved by *AT* (88.3 %), followed by *AV* (78.8 %) and *BV* (71.1 %). Whereas finding a "city scene" achieved high recognition rates throughout all agents, commands leading to finding a "mountain" scene were well recognized by both advanced agents (*AV*: 88.9 %, *AT*: 96.9 %), but very poorly by *BV* (55.1 %). In contrast, *BV* achieved a higher recognition rate for the task of finding a "warm forest scene" (81.2 %) than the advanced agents (*AV*: 71.0 %, *AT*: 74.6 %). In addition, *AV* had difficulties recognizing commands to find a "bright and blue scene", compared to *AT* (68.4 % vs. 89.6 %).

6. Discussion

6.1. Smartphone Application

Study results showed that the *Smartphone Application* as the simplest and most familiar system was also the most favorable system. It achieved the highest SUS score, the best overall ranking, and was the system that exceeded users' expectations the most. They liked the "simple and clear graphical user interface" (S06) that facilitated quick and efficient interactions, and especially appreciated the visual preview of scenes (S11) (see also Table 4). On the other hand, the lack of autonomy required users to resolve the tasks and map their requests to scene properties all by themselves. Accordingly, they reported that using the smartphone application felt more like "controlling a fancy light switch" (S08) rather than having an actual interaction, and made the office space appear "inanimate and lonely" (S08). On average, the smartphone application had the lowest task completion time. It outperformed other systems for tasks that could directly be mapped to the sliders of the user interface, such as finding a bright and blue or a warm forest scene. However, users criticized that adjusting the sliders can be time consuming and "not as intuitive for more abstract tasks" (S24), like finding a relaxing scene. Additionally, in an office scenario, users might be more susceptible to distraction by the smartphone, causing disruption in their workflow as they would have to switch from the computer to the smartphone and back [1]. However, whether regular users of the smart office prototype might even be willing to accept this potentially more distracting interface in return for a simpler and quicker way of interaction still needs to be analyzed in further detail.

Table 4. Strengths and weaknesses for each agent.

	Strengths	Weaknesses
Smartphone Application	- Quick, familiar interaction - Predictable actions - Visual scene preview - Efficient filtering of scenes	- No hands-free interaction - No interaction with room - No autonomy - Disrupts workflow easily
Basic Voice Agent	- Hands-free interaction - Manageable feature range, no overtaxing of user	- Rigid, inanimate, time-consuming interaction - No intelligence or context-awareness - Strong familiarity with system required
Advanced Voice Agent	- Hands-free interaction - Engaging, conversational - Abstract scene description possible	- Bad smalltalk handling - Wide gulf of user expectation and experience - False recognition increases with request complexity
Advanced Text Agent	- Fast interaction, especially in office scenario - Engaging, conversational - Abstract scene description possible - Narrow gulf of user expectation and experience	- No hands-free interaction - Bad smalltalk handling - False recognition increases with request complexity

6.2. Basic Voice Agent

The *Basic Voice Agent* created the impression of being very "rigid and time consuming" (S28), often reminding users of "talking to a customer service hotline" (S08). Additional problems with voice recognition, causing users to repeat the same commands over and over, made them feel uncomfortable when interacting with the agent (S16). Combined with a lack of system intelligence and context-awareness, it led to the lowest overall usability, and also the lowest ratings of intelligence and engagement. However, some users appreciated the limited range of features. In their opinion, it made the basic agent appear structured and led to predictable results—as long as the voice recognition worked correctly. Users were able to accomplish some tasks equally fast or even faster with the basic agent than with the advanced agents. For example, for the "warm forest" task, they simply looked for forest scenes in general and then iterated through them using the "next" command until they found one with warm color temperature.

6.3. Advanced Voice Agent

Users reported that the *Advanced Voice Agent* was very friendly and cooperative in finding a matching scene. The possibility of transforming the workspace using a more open command list, such as providing a high level description of scenes, or by letting the agent autonomously recommend scenes, was appreciated throughout all users. The agent tried to conduct a normal conversation by referring to users by their names, using a more informal language, and offering a broader variety of text responses (S18). In the questionnaires, users referred to the agent as "she" instead of "it" because it appeared emotional and engaging, and even made users users laugh occasionally (S25). However, the SUS score yields that the advanced voice agent was rated worse than the average score, which was mainly due to the voice recognition. It was reported as the main bottleneck that impairs the user experience and creates a gulf between user expectations towards a seemingly intelligent system. User experience was compromised by the system being unable to correctly recognize commands such as "find a bright and blue scene", which was often falsely recognized as "find a bride and blue scene". As results indicate, native speakers were more critical towards voice recognition because

their judgement appeared to be more determined by correct voice recognition and a smooth, natural conversation than by the actual range of features. Therefore, they rated this agent more similar to the *Basic Voice Agent*, mentioning that the lack of voice recognition accuracy made it feel dumb. In contrast, non-native speakers noticed a clear difference between both voice agents. Users also reported that the interaction over voice could be time consuming because it sometimes took too long to get a response from the agent, especially when it did not understand the command.

6.4. Advanced Text Agent

The *Advanced Text Agent* achieved the second highest overall usability and was the only agent with a SUS score higher than the average score. Compared to the *Advanced Voice Agent*, the interaction with this agent was found to be smoother and more straightforward as the agent instantly provided a response over text. Similarly, it was on average rated as being more engaging, especially among native speakers, indicating that the use of emojis was an effective compensation for the missing voice interface.

However, the *Advanced Text Agent* was the only agent that achieved higher perceived levels of engagement among native speakers compared to non-native speakers. Replacing the voice interface by a text-based interface, and, thus, enabling smoother conversation by preventing occasionally faulty voice recognition, made the *Advanced Text Agent* appear far more intelligent or engaging for native-speaker. This further supports the conclusion that native speakers are more critical towards voice recognition, which heavily impairs user experience if not working reliably. In contrast, non-native speakers were not as critical of voice recognition as native speaker (indicated by comparable levels of both intelligence and engagement for the two advanced agents among non-native speakers) even though some non-native speakers reported that the voice agent apparently had trouble understanding their accent (e.g., S03, S20, S29). Our findings are, to some extent, in contrast to the work of Pyae and Scifleet [49] who reported that native speakers had a better overall user experience with smart speakers, such as Google Home, than non-native speakers. However, their study was focused on exploring the general features of the speaker rather than evaluating a custom application for a particular use case. Future work, nonetheless, needs to follow-up on these results in further detail.

Furthermore, we observed that users used different wordings to communicate with the two advanced agents. For instance, when trying to find a relaxing scene, they asked the voice agent "Can you find a scene that makes me relax?", whereas they used more informal and shorter commands for the text agent, like "Relaxing scene" or even just "Relax". Users reported that it felt strange talking to an agent in such an informal manner, but it did not when they texted the system, since it is also usual for them to text their friends in the same way. However, participants emphasized that they were confused by both advanced agents of being conversational on the one side, but offering only limited understanding for small talk and filler language.

On average, both *AV* and *AT* required less interactions than the smartphone application and can more easily be integrated into the workflow of office workers. They can interact with the agent while performing other tasks in the meantime. Moreover, the radius of interaction at the workplace is usually more restricted to desk, compared to a smart home scenario. Therefore, it might be easier to interact with a text agent over the computer by simply switching to the agent's conversation window rather than having to speak out loud to a voice agent. For future applications, a hybrid agent, offering the possibility to interact via voice and text depending on the current situation, might considerably improve user experience as well as productivity in the smart office. Additionally, moving away from general-purpose voice and text agents to a more domain-specific for the office context will be important for increasing the acceptance [56].

7. Conclusions and Outlook

We introduced smart agents, differing in system intelligence and input modality, for the interaction with an ambient adaptive environment. Using such agents, users could dynamically change the

appearance of their workspace either by switching to particular scenes, by describing scenes in an abstract way, or by letting the agent recommend scenes based on a certain context. We designed and conducted an experiment to evaluate the different agents and to find out which one is the best for the smart office context. Our results show that the *Smartphone Application* as the most familiar user interface is also the system that achieved the highest usability and the best overall ranking. It was followed by the *Advanced Text Agent* and the *Advanced Voice Agent*, which were both found to be very friendly and engaging conversational agents that facilitate finding the right scenes. However, especially the *Advanced Voice Agent* created a clearly different perception among native and non-native speakers. For their judgement, non-native speakers considered more the range of features, whereas native speakers were more influenced by correct voice recognition and a smooth, natural conversation. Future improvements in voice user interfaces, especially with an enhanced ability to understand smalltalk and filler language, could make significant improvements in closing the gulf of evaluation between user expectation and experience.

In general, our experiment shows that all agents have their strengths and weaknesses, with the *Advanced Text Agent* offering a well appreciated trade-off between a quick and easy interaction, a natural, engaging, and entertaining conversation, and an open command list with the possibility to let the agents recommend scenes. Future work should, therefore, synthesize our findings and combine different interaction modalities to build a flexible, multimodal agent for the smart office domain—similar to the Jarvis AI developed by Mark Zuckerberg for his home [57]. The agent would consist of a chatbot as core part, enhanced with visual scene previews integrated into the conversation feed. Additionally, hands-free interaction over voice commands should be possible, for instance if the occupant is entering or leaving the office or performs a hands-demanding task. Evaluating the usage (or avoidance) behavior of the different modalities of such a multimodal agent in a longitudinal study would allow to better identify the individual strengths and weaknesses, and would reveal which modality lags behind the others, and which would need to be better adapted for the particular use case in an adaptive smart office scenario. Finally, we are confident that such a multimodal agent will eventually be beneficial for future smart office systems by reducing work and information overload and foster creativity, productivity, and relaxation.

Author Contributions: Conceptualization: R.R., N.Z., B.M.E., J.A.P.; methodology: R.R., N.Z., B.M.E., J.A.P.; software and data curation: R.R.; validation and formal analysis: R.R., N.Z.; writing—original draft preparation, R.R.; writing—review and editing, R.R., N.Z., B.M.E., J.A.P.; visualization: R.R.; supervision: B.M.E., J.A.P.; funding acquisition: B.M.E., J.A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a fellowship within the FITweltweit program of the German Academic Exchange Service (DAAD). Bjoern M. Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship programme (grant number ES 434/8-1).

Acknowledgments: We thank all subjects that participated in this study.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANOVA	Analysis of variance
API	Application Programming Interface
AT	Advanced Text Agent
AV	Advanced Voice Agent
BV	Basic Voice Agent
M ± SD	Mean ± Standard Deviation
SP	Smartphone Application
SUS	System Usability Scale

References

1. Smith, A. U.S. Smartphone Use in 2015. Available online: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015 (accessed on 15 April 2020).
2. Andrews, S.; Ellis, D.A.; Shaw, H.; Piwek, L. Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use. *PLoS ONE* **2015**, *10*, e0139004. [CrossRef] [PubMed]
3. Google. Google Assistant, Your Own Personal Google. Available online: https://assistant.google.com/ (accessed on 11 April 2020).
4. Geller, T. Talking to machines. *Commun. ACM* **2012**, *55*, 14–16. [CrossRef]
5. Brennan, S. Conversation as Direct Manipulation: An Iconoclastic View. In *The Art of Human-Computer Interface Design*; Laurel, B., Mountford, S.J., Eds.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1990.
6. Mozer, M.C. An Intelligent Environment must be adaptive. *IEEE Intell. Syst.* **1999**, *14*, 11–13. [CrossRef]
7. Kidd, C.D.; Orr, R.; Abowd, G.D.; Atkeson, C.G.; Essa, I.A.; MacIntyre, B.; Mynatt, E.; Starner, T.E.; Newstetter, W. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In Proceedings of the International Workshop on Cooperative Buildings, Pittsburgh, PA, USA, 1–2 October 1999; pp. 191–198.
8. Zhao, N.; Azaria, A.; Paradiso, J.A. Mediated Atmospheres: A Multimodal Mediated Work Environment. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 1–23. [CrossRef]
9. Gensler Research Institute. 2013 U.S. Workplace Survey Key Findings. Available online: https://www.gensler.com/research-insight/gensler-research-institute/the-2013-us-workplace-survey-1 (accessed on 25 April 2020).
10. Maes, P. Agents that reduce work and information overload. *Commun. ACM* **1994**, *37*, 30–40. [CrossRef]
11. Luger, E.; Sellen, A. Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7 May 2016; pp. 5286–5297.
12. Rossi, M.; Pandharipande, A.; Caicedo, D.; Schenato, L.; Cenedese, A. Personal lighting control with occupancy and daylight adaptation. *Energy Build.* **2015**, *105*, 263–272. [CrossRef]
13. Chen, N.H.; Nawyn, J.; Thompson, M.; Gibbs, J.; Larson, K. Context-aware tunable office lighting application and user response. In *SPIE Optical Engineering+ Applications*; International Society for Optics and Photonics: Washington, DC, USA, 2013; p. 883507.
14. Aldrich, M.; Zhao, N.; Paradiso, J. Energy efficient control of polychromatic solid state lighting using a sensor network. In *SPIE Optical Engineering+ Applications*; International Society for Optics and Photonics: Washington, DC, USA, 2010; p. 778408.
15. Kalyanam, R.; Hoffmann, S. Visual and thermal comfort with electrochromic glass using an adaptive control strategy. In Proceedings of the 15th International Radiance Workshop (IRW), Padua, Italy, 30 August 2016.
16. Feldmeier, M.; Paradiso, J.A. Personalized HVAC control system. In Proceedings of the 2010 Internet of Things (IOT), 29 November–1 December 2010; pp. 1–8.
17. Kainulainen, A.; Turunen, M.; Hakulinen, J.; Salonen, E.P.; Prusi, P. A Speech-based and Auditory Ubiquitous Office Environment. In Proceedings of the 10th International Conference on Speech and Computer, Patras, Greece, 17–19 October 2005.
18. Raskar, R.; Welch, G.; Cutts, M.; Lake, A.; Stesin, L.; Fuchs, H. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, Anaheim, CA, USA, 21–25 July 1998; pp. 179–188.
19. Tomitsch, M.; Grechenig, T.; Moere, A.V.; Renan, S. Information Sky: Exploring the Visualization of Information on Architectural Ceilings. In Proceedings of the 12th International Conference on Information Visualisation (IV '08), London, UK, 8–11 July 2008; pp. 100–105.
20. Johanson, B.; Fox, A.; Winograd, T. The Interactive Workspaces project: Experiences with ubiquitous computing rooms. *IEEE Pervasive Comput.* **2002**, *1*, 67–74. [CrossRef]
21. Pejsa, T.; Kantor, J.; Benko, H.; Ofek, E.; Wilson, A. Room2Room: Enabling life-size telepresence in a projected augmented reality environment. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, CA, USA, 27 February–2 March 2016; pp. 1716–1725.

22. Schnädelbach, H.; Glover, K.; Irune, A.A. ExoBuilding: Breathing Life into Architecture. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Reykjavik, Iceland, 16–20 October 2010; pp. 442–451.
23. Ishii, H.; Wisneski, C.; Brave, S.; Dahley, A.; Gorbet, M.; Ullmer, B.; Yarin, P. ambientROOM: Integrating Ambient Media with Architectural Space. In Proceedings of the CHI 98 Conference Summary on Human Factors in Computing Systems, Los Angeles, CA, USA, 18–23 April 1998; pp. 173–174.
24. Mozer, M.C. The neural network house: An environment that adapts to its inhabitants. In Proceedings of the AAAI Spring Symposium on Intelligent Environments, Palo Alto, CA, USA, 23–25 March 1998; Volume 58.
25. Mozer, M.C.; Dodier, R.; Miller, D.; Anderson, M.; Anderson, J.; Bertini, D.; Bronder, M.; Colagrosso, M.; Cruickshank, R.; Daugherty, B.; et al. The adaptive house. In Proceedings of the IEE Seminar on Intelligent Building Environments, Colchester, UK, 28 June 2005; Volume 11059.
26. Mennicken, S.; Vermeulen, J.; Huang, E.M. From today's augmented houses to tomorrow's smart homes: New directions for home automation research. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13–17 September 2014; pp. 105–115.
27. Cook, D.J.; Youngblood, M.; Heierman, E.O.; Gopalratnam, K.; Rao, S.; Litvin, A.; Khawaja, F. MavHome: An Agent-Based Smart Home. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, Fort Worth, TX, USA, 23–26 March 2003; pp. 521–524.
28. Abras, S.; Ploix, S.; Pesty, S.; Jacomino, M. A Multi-agent Home Automation System for Power Management. In *Informatics in Control Automation and Robotics*; Springer: Berlin, Germany, 2008; pp. 59–68.
29. Alan, A.T.; Costanza, E.; Ramchurn, S.D.; Fischer, J.; Rodden, T.; Jennings, N.R. Tariff Agent: Interacting with a Future Smart Energy System at Home. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **2016**, *23*, 25. [CrossRef]
30. Danninger, M.; Stiefelhagen, R. A context-aware virtual secretary in a smart office environment. In Proceedings of the 16th ACM International Conference on Multimedia, Vancouver, BC, USA, 27–31 October 2008; pp. 529–538.
31. Bagci, F.; Petzold, J.; Trumler, W.; Ungerer, T. Ubiquitous mobile agent system in a P2P-network. In Proceedings of the Fifth Annual Conference on Ubiquitous Computing, Seattle, WA, USA, 12 October 2003; pp. 12–15.
32. Hossain, M.A.; Shirehjini, A.A.N.; Alghamdi, A.S.; El Saddik, A. Adaptive interaction support in ambient-aware environments based on quality of context information. *Multimed. Tools Appl.* **2013**, *67*, 409–432. [CrossRef]
33. Straßmann, C.; von der Pütten, A.R.; Yaghoubzadeh, R.; Kaminski, R.; Krämer, N. The Effect of an Intelligent Virtual Agent's Nonverbal Behavior with Regard to Dominance and Cooperativity. In Proceedings of the International Conference on Intelligent Virtual Agents, Los Angeles, CA, USA, 20–23 September 2016; pp. 15–28.
34. Mennicken, S.; Zihler, O.; Juldaschewa, F.; Molnar, V.; Aggeler, D.; Huang, E.M. It's like living with a friendly stranger: Perceptions of personality traits in a smart home. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 120–131.
35. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *59*, 433–460. [CrossRef]
36. Weizenbaum, J. ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]
37. D'Haro, L.F.; Kim, S.; Yeo, K.H.; Jiang, R.; Niculescu, A.I.; Banchs, R.E.; Li, H. CLARA: A Multifunctional Virtual Agent for Conference Support and Touristic Information. In *Natural Language Dialog Systems and Intelligent Assistants*; Springer: Berlin, Germany, 2015; pp. 233–239.
38. Gnewuch, U.; Morana, S.; Maedche, A. Towards Designing Cooperative and Social Conversational Agents for Customer Service. In Proceedings of the Thirty Eigth International Conference on Information Systems, Seoul, Korea, 10–13 December 2017; pp. 1–13.
39. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. *Commun. ACM* **2016**, *59*, 96–104. [CrossRef]
40. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef]
41. Montenegro, J.L.Z.; da Costa, C.A.; da Rosa Righi, R. Survey of conversational agents in health. *Expert Syst. Appl.* **2019**, *129*, 56–67. [CrossRef]
42. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied conversational agents in clinical psychology: A scoping review. *J. Med. Internet Res.* **2017**, *19*. [CrossRef] [PubMed]

43. Chu, Z.; Gianvecchio, S.; Wang, H.; Jajodia, S. Who is tweeting on Twitter: Human, bot, or cyborg? In Proceedings of the 26th Annual Computer Security Applications Conference, Austin, TX, USA, 6–10 December 2010; pp. 21–30.
44. Subrahmanian, V.; Azaria, A.; Durst, S.; Kagan, V.; Galstyan, A.; Lerman, K.; Zhu, L.; Ferrara, E.; Flammini, A.; Menczer, F. The DARPA Twitter bot challenge. *Computer* **2016**, *49*, 38–46. [CrossRef]
45. Schaffer, S.; Schleicher, R.; Möller, S. Modeling input modality choice in mobile graphical and speech interfaces. *Int. J. Hum. Comput. Stud.* **2015**, *75*, 21–34. [CrossRef]
46. Weiß, B.; Möller, S.; Schulz, M. Modality Preferences of Different User Groups. In Proceedings of the Fifth International Conference on Advances in Computer-Human Interactions (ACHI 2012), Valencia, Spain, 30 January–4 February 2012.
47. Silva, S.; Almeida, N.; Pereira, C.; Martins, A.I.; Rosa, A.F.; Oliveira e Silva, M.; Teixeira, A. Design and Development of Multimodal Applications: A Vision on Key Issues and Methods. In *Universal Access in Human-Computer Interaction, Proceedings of the Today's Technologies: 9th International Conference, UAHCI 2015, Los Angeles, CA, USA, 2–7 August 2015*; Springer International Publishing: Cham, Switzerland, 2015.
48. Kühnel, C.; Westermann, T.; Weiß, B.; Möller, S. Evaluating Multimodal Systems: A Comparison of Established Questionnaires and Interaction Parameters. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction (NordiCHI 2010), Reykjavik, Iceland, 16–20 October 2010; pp. 286–294.
49. Pyae, A.; Scifleet, P. Investigating the Role of User's English Language Proficiency in Using a Voice User Interface: A Case of Google Home Smart Speaker. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–6.
50. Sivasubramanian, S. Amazon dynamoDB: A seamlessly scalable non-relational database service. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AZ, USA, 20–24 May 2012; pp. 729–730.
51. Howarth, E.; Hoffman, M.S. A multidimensional approach to the relationship between mood and weather. *Br. J. Psychol.* **1984**, *75*, 15–23. [CrossRef]
52. Luminous, S. The Circadian Rhythm and Color Temperature. Available online: https://sigmaluminous.com/the-circadian-rhythm-and-color-temperature (accessed on 15 April 2020).
53. Duffy, J.F.; Czeisler, C.A. Effect of light on human circadian physiology. *Sleep Med. Clin.* **2009**, *4*, 165–177. [CrossRef]
54. Brooke, J. SUS—A quick and dirty usability scale. *Usability Eval. Ind.* **1996**, *189*, 4–7.
55. Hone, K.S.; Graham, R. Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Nat. Lang. Eng.* **2000**, *6*, 287–303. [CrossRef]
56. Mennicken, S.; Brillman, R.; Thom, J.; Cramer, H. Challenges and Methods in Design of Domain-Specific Voice Assistants. In *2018 AAAI Spring Symposium Series*; Association for the Advancement of Artificial Intelligence: Park, CA, USA, 2018; pp. 431–435.
57. Zuckerberg, M. Building Jarvis. Available online: https://www.facebook.com/notes/mark-zuckerberg/building-jarvis/10154361492931634 (accessed on 10 April 2020).