Serious Games for Training Social Skills in Job Interviews

Patrick Gebhard, Tanja Schneeberger, Elisabeth André, Tobias Baur, Ionut Damian, Gregor Mehlmann, Cornelius König, and Markus Langer

***Abstract*—In this paper, we focus on experience-based role play with virtual agents to provide young adults at the risk of exclusion with social skill training. We present a scenario-based serious game simulation platform. It comes with a social signal interpretation component, a scripted and autonomous agent dialog and social interaction behavior model, and an engine for 3-D rendering of lifelike virtual social agents in a virtual environment. We show how two training systems developed on the basis of this simulation platform can be used to educate people in showing appropriate socioemotive reactions in job interviews. Furthermore, we give an overview of four conducted studies investigating the effect of the agents' portrayed personality and the appearance of the environment on the players' perception of the characters and the learning experience.**

***Index Terms*—.**

Fig. 1. User interacting with TARDIS. Paperboard cards give hints on how to behave for each phase of a job interview.

I. Introduction

PEDAGOGICAL role play with virtual agents offers great promise for social skill training. It provides learners with a realistic, but safe environment that enables them to train specific verbal and nonverbal behaviors in order to adapt to socially challenging situations. At the same time, learners benefit from the gamelike environment, which increases not only their enjoyment and motivation but also enables them to take a step back from the environment and think about their behavior if necessary.

In this paper, we will present a scenario-based serious game simulation platform that supports social training and coaching in the context of job interviews. The game simulation platform has been developed in the TARDIS project [1] and further extended in the EmpaT project [2]. The platform includes technology to detect the users' emotions and social attitudes in real time through voice, gestures, and facial expressions during the interaction with a virtual agent as a job interviewer. To achieve their pedagogical goals, TARDIS and EmpaT need to expose the players to situations in the learning environment that evoke similar reactions in them as real job interviews. They require a high demand for computational intelligence and perceptual skills in order to understand the player's socioemotional reactions and optimally adapt the pace of learning.

In TARDIS, users were able to interact with a virtual recruiter that responded to their paraverbal and nonverbal behaviors (see Fig. 1). However, users were not immersed in the physical setting in which the job interview took place (e.g., the building and the room style, the employees, or the specific atmospheric setup). Furthermore, the TARDIS users' experience is limited to the job interview setup, in which the user sits in front of the virtual job recruiter at a desk.

EmpaT embeds the job interview into a virtual environment that comes with a virtual personal assistant who explains every step of the job interview experience. Moreover, the virtual environment allows simulating various challenges that come along with job interviews, as that users may navigate through to find the room where the actual job interview will take place (see Fig. 2). On their way to the interview, users arrive to the reception desk asking for the job interview appointment and wait until they are called for the interview in the nearby lobby. In

Manuscript received December 30, 2016; revised August 31, 2017; accepted January 13, 2018. Date of publication; date of current version. This work was supported in part by the German Ministry of Education and Research (BMBF) within the EmpaT project (funding code 16SV7229K) and in part by the European Commission within FP7-ICT-2011-7 (project TARDIS, Grant Agreement 288578). *(Corresponding author: Patrick Gebhard.)*

P. Gebhard and T. Schneeberger are with the German Research Centre for Artificial Intelligence, Saarbrücken 66123, Germany (e-mail: gebhard@dfki.de; tanja.schneeberger@dfki.de).

E. André, T. Baur, I. Damian, and G. Mehlmann are with the Augsburg University, Augsburg 86159, Germany (e-mail: andre@informatik.uni-augsburg.de; baur@hcm-lab.de; damian@hcm-lab.de; gregor.mehlmann@informatik.uni-augsburg.de).

C. König and M. Langer are with the Saarland University, Saarbrücken 66123, Germany (e-mail: ckoenig@mx.uni-saarland.de; markus.langer@uni-saarland.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TG.2018.2808525

Fig. 2. Company building in EmpaT, in which the interview takes place.

the waiting phase, users can observe the daily routine of the simulated employees. The EmpaT system allows confronting users with situations that might increase their uneasiness, for example, when having to ask unfriendly personnel for directions or in case of interruptions during the actual job interview. Thus, EmpaT enables a more comprehensive experience that includes all phases of a job interview from entering to leaving the building of the company where the job interview takes place.

In the following, we will first discuss related work on the use of computer-enhanced role play for social coaching. After that, we will analyze elements of game design that may have an impact on the achievement of pedagogical goals in social coaching. We then present the serious game simulation platform that supports social learning in the context of job interviews. Finally, we present four studies we conducted to investigate the impact of serious games for social skill training and the influence of the agents' behaviors and the physical environment on the players' perception of the agents and the learning experience.

II. Related Work

Computerized social skill training tools have seen rapid evolution in recent years due to advances in the areas of social signal processing as well as improvements in the audio-visual rendering of virtual agents. Such tools are meant to complement or even substitute traditional training approaches.

A variety of serious games employs role play with virtual agents that foster reflection about socioemotional interactions. An example includes the anti-bullying Game FearNot! that has been developed within the project eCircus [3]. The project investigates how social learning may be enhanced through interactive role play with virtual agents that establish empathetic relationships with the learners. It creates interactive stories in a virtual school with embodied conversational agents in the role of bullies, helpers, victims, etc. The children run through various bullying episodes, interact with the virtual agents after each episode, and provide advice to them. The benefit of educational role plays of this kind lies in the fact that they promote reflective thinking. Results of a conducted evaluation [4] showed that the system had a positive effect on the children's abilities to cope with bullying.

Role play with virtual agents has also been a popular approach to educate users in cultural sensitivity. Employing role play with virtual agents that represent different cultures, users are supposed to develop a better understanding of other cultures. Eventually, the users are expected to develop intercultural empathy and reduce their negative attitude toward other cultures. An example of such a system has been developed within the eCute project: The objective of MIXER (moderating interactions for cross-cultural empathic relationships)[1] is to enable users to experience emotions that are usually elicited during interactions of members of a different group [5]. To this end, children are confronted with scenarios in which virtual agents appear to violate previously introduced rules in a game scenario. Such a situation leads inevitably to frustration and negative attitudes toward members of the other group. By interacting with MIXED, children are expected to learn to reflect about behaviors of other groups and reconsider potentially existing prejudices against them. The setting was inspired by the card-game BARNGA, which has been successfully used for cultural training of adults [6]. Other than expected by the authors, the MIXER game did not foster cultural awareness in children in a pilot study. The authors assumed that the learning objectives MIXER was designed to meet were not appropriate for the age group that was not able to cope with the negative rule-clash-based conflict.

While the above-described systems analyze the user's verbal and nonverbal behaviors for the purpose of the interaction, their primary objective is to help users cope with socially challenging situations. They do not aim at teaching users appropriate socioemotional communication skills directly.

Within the project ASD-Inclusion [7], techniques for the recognition of human socioemotional behaviors have been employed to help children who have autism to improve their socioemotional communication skills. A game platform with virtual agents has been developed that enables children to learn how emotions can be expressed and recognized via gestures, facial, and vocal expressions in a virtual game world. A requirement analysis revealed the need to incorporate an appropriate incentive system to keep children engaged. Therefore, the authors implemented a monetary system which rewarded children with virtual money for improved performance from which they could buy items for their avatars.

Furthermore, social signal processing techniques have been employed to automatically record and analyze the learner's social and emotional signals, whereas virtual agents are employed to simulate various social situations, such as social gatherings [8] or public speeches [9]. Similar to our work is a job interview simulation with a virtual agent by Hoque *et al.* [10]. They explored the impact of the job interview training environment on MIT students and concluded that students who used the system to train, experienced a larger performance increase than students who used conventional methods. These results are encouraging for our research. However, while Hoque *et al.* recruited MIT students as participants, our target group are job-seeking young

[1] http://ecute.eu/mixer/

people who have been categorized as being at risk of exclusion. Furthermore, they did not explicitly incorporate elements from games to increase the players' motivation.

A number of studies reveal the positive effects of gamelike environments for social coaching. However, the research conducted in the eCute project also points out difficulties in designing a gamelike environment that achieves particular pedagogical goals. Overall, there is still a lack of knowledge on the relationship between specific game attributes and learning outcomes. In the next section, we use the taxonomy by Bedwell and colleagues [11] as a starting point for the analysis of game attributes in TARDIS and EmpaT.

III. Game Experience

To support social coaching in TARDIS and EmpaT, we incorporated elements from serious games for which we hypothesized a positive effect on learning. To this end, we consulted the work by Wilson *et al.* [12] as well as Bedwell *et al.* [11] who identified eight categories of game attributes designers should be especially aware of when developing gamified environments: action language, assessment, conflict/challenge, control, environment, game fiction, human interaction, immersion, and rules/goals. In the remainder of this section, we take a closer look upon seven of these game attribute categories (we will not include human interaction, as there is no human interaction in the two job interview training games) and describe to what extent they have been taken into account during the design of the job interview games in TARDIS and EmpaT.

A. Nonverbal and Paraverbal Behavior as an "Action Language"

Action language defines the way how users interact with the game (e.g., by using a joystick or a keyboard). It is an important aspect to consider when designing gamified environments as the mode of interaction may have a strong influence on the learning outcome [12]. In commercial computer games, the action language employed to communicate with the game represents a well-defined mapping between commands to be input by the user and actions to be executed by the game. Unlike commercial games, TARDIS and EmpaT rely on natural forms of interaction with focus on paraverbal and nonverbal behavior to which the interview agents react in a believable manner.

This form of interaction poses particular challenges to the design of the interaction. Due to deficiencies of current technology to process natural language input, effective strategies had to be found to support a consistent and coherent conversational flow. Based on an evaluation of Façade, a gamelike interactive storytelling scenario with conversational agents, Mehta *et al.* [13] came up with a number of guidelines and recommendations for dialogue design in gamelike environments, such as avoiding shallow confirmations of user input and supporting the user's abilities to make sense of recognition flaws. Both in TARDIS and in EmpaT, the user is supposed to play a role that is in accordance by the learning goals. To support a smooth conversational flow, the virtual agents provide explicit interaction prompts. That is the agents are modeled in a way that they are requesting specific information from the user. This way, the user knows what kind of input is required and learns at the same time which questions are typically asked in a job interview. As long as the user follows the rules of the game, there is no need to conduct a deep semantic analysis of the user's utterances even though some simple form of keyword spotting has shown beneficial. Due to the design of the scenario, failures of the natural language understanding technologies could be interpreted as communication issues that typically arise in job interviews. For example, a virtual job interviewer shifting to another topic due to natural language understanding problems may still provide a compelling performance, for example, by indicating boredom of the previous topic. Text-based input would facilitate the analysis of natural language input significantly. However, this option had to be discarded in our case. First, text-based input would break the illusion of a realistic job interview. Second, users are expected to acquire appropriate paraverbal and nonverbal behaviors that have to be synchronized with their speech. Consequently, the game environment should enable them to practice these behaviors.

B. Assessment Through Social Sensing

Assessment refers to the feedback given to the user on their progress [14]. In order to keep users motivated, it is essential to provide feedback to them on how well they are doing so far and how advanced they are regarding specific goals [11]. In a social setting with virtual agents, direct feedback can be given naturally by the agents' nonverbal and verbal cues. However, users might not always understand such implicit cues. Learning to read somebody's body language could be the topic of a serious game on its own, but would distract from the actual learning goals here. In order to increase the agents' believability in TARDIS and EmpaT, they respond immediately to the user's input by appropriate nonverbal and verbal cues. However, we also incorporated more explicit feedback in TARDIS and EmpaT that helps users improve their behavior in subsequent interactions.

In TARDIS, we implemented a reward system that remunerates users after execution of successful actions. To encourage adequate behaviors, the system scores the users' performance and rewards him or her with points if he or she behaves in compliance with behaviors specified on a game card (see Fig. 1). A score for the user's behavior is computed in real time during the interaction by using sensing devices to recognize social cues, such as a smile or crossed arms. Providing feedback on social behavior is an ambitious task due to the high amount of subjectivity and lack of transparency. For example, it may be counterproductive to tell the user that he or she appears disengaged without giving him or her the reasons for such an assessment. Therefore, TARDIS offers additional feedback to users in a debriefing phase through a graphical user interface that highlights social cues that contributed to the system's assessment of the user's behavior (see Section IV-D).

In EmpaT, we are currently exploring possibilities of giving users continuous feedback on their behavior and progress. The challenge consists in providing such feedback without disturbing the flow of the game. Currently, we are investigating the use of signal lights to give feedback on paraverbal and nonverbal behavior dynamically and in real time. For example, the signal light for eye contact would turn red if someone is not keeping eye contact with the interviewer for a predefined ratio of time, but the signal light would adapt dynamically and turn green again if the participant succeeds in keeping eye contact for longer than the above-mentioned ratio of time. Furthermore, we are studying immediate reactions of the virtual interview agent to the users' behavior, such as exhorting users if they interrupt the virtual agent during its speech. This kind of assessment raises awareness of how to behave during job interviews and enable them to learn how to apply nonverbal behavior adequately. Furthermore, positive feedback improves the users' self-efficacy and enhances their motivation to keep on training social skills behaviors.

C. Different Levels of Conflict/Challenge

Adding conflict/challenge leads to difficulties and problems within the game that need to be solved, as well as to uncertainties enhancing the tension. For instance, random events like employees coming into the interview room and disturbing the interaction can add unforeseeable aspects. Another example would be that participants can be confronted with job interview questions of varying difficulty enhancing replayability. Thus, conflict/challenge is a driving force within the game that keeps the users motivated to proceed [11], [15]. It is important to note that it is crucial to define difficulty levels carefully, so the game is sufficiently challenging, but not too difficult [12].

Within TARDIS and EmpaT, we implemented various levels of difficulty offering a challenging experience for users with different levels of job interview experience.

TARDIS makes use of one virtual agent with different social behavior profiles, understanding and demanding, which consequently influence the level of difficulty of the simulation as well as the impact on the user.

In EmpaT, job interviews are performed by one out of two virtual interviewers of different age: a young and middle-aged male, and a 50-years old female (see Fig. 3, center and right-hand sides) reflecting experience and status of the agent [16]. Furthermore, these agents express different nonverbal and verbal behaviors which portray the agents' personality (understanding, demanding, and neutral) [17]. Depending on their personality profile, these agents evoke emotions in the user that are experienced in real job interviews and thus enhance the realism of the simulation (see Section V). Also, the EmpaT realization provides users with an understanding personal assistant that guides the user through the interview experience (see Fig. 3, left-hand side).

In addition to increasing the level of difficulty by agents representing a higher status, EmpaT introduces critical events in the job interview. For instance, in an entry level job interview, there is a young interview agent in casual clothing behaving in amiable manner and asking easy and common interview questions. In comparison, at a higher level, the age and appearance of the interview agent reflect a more experienced member of the organization or even the leader of the company. Questions in the higher level job interview are less common or even provoking. Thus, interviewees have to adapt to the enhanced degree of difficulty through different behavior. Also, random events can be added. For example, another virtual agent might enter the room or the interviewer might make a challenging comment on the user's behavior. This way, the game can be modulated to create tension and stress in the users similarly to a real job interview situation, thus enhancing the realism of the simulation. Providing challenges to the users can lead to reduced anxiety in real job interview situations and improved self-efficacy because the users already have experienced similar situations in the training game. Moreover, customizable difficulty and random events enhance replayability, further increasing exposure to the training environment.

Fig. 3. Virtual 3-D environment (VRE) social agents.

D. Guided Control

Control describes how much users can influence the game by their actions [11], [15]. A high level of control can positively impact the users' experience, but it can also be detrimental if users get lost within the environment [11]. Within the EmpaT job interview training, the user can walk around freely to explore the virtual environment. However, at some point, the user will be led to the meeting room by the virtual interviewer.

When designing the dialog with the virtual interviewer, the question arises of how much control should be given to the user. A mixed-initiative dialog gives more freedom to the user. However, it also requires more sophisticated language understanding capabilities than system-initiative dialog. In [18], we compared the system-initiative dialog with mixed-initiative dialog in a soap-opera-like game environment that included a text input interface to enable users to communicate with virtual agents. The users preferred the mixed-initiative dialog over the system-initiative dialogue even though the mixed-initiative dialog was less robust. Apparently, the experiential advantages of the mixed-initiative dialog compensated for the lower amount of accuracy in natural language understanding.

TARDIS and EmpaT rely on a speech-based input which comes with even greater challenges than a text-based input. Therefore, we decided to implement the less demanding option of system-initiative dialog in order to ensure a smooth flow of dialogue. This interaction style appears to match the situation of a job interview well where the applicants are not expected to take over control. Furthermore, the system-initiative dialog still gives autonomy to the users. During the actual interview, users can focus on the main aspects of the simulation: the questions the interviewer asks, their answers, and their paraverbal and nonverbal behavior—still leaving a high level of control to users through speech and body movement. Thus, the simulation and its outcomes depend on users' own actions. This setup enhances realism and gives users the opportunity to experiment with their nonverbal behavior and learn about consequences.

E. Realistic Environment

The environment defines where users find themselves in the game and how they see the world [11]. In EmpaT, users see the world in first person view as they walk through a realistic office building. The entrance hall of the company building has a reception desk, where users are welcomed by a virtual agent, a waiting room where users wait to be picked up by the interview agent, and various rooms where the interview can be conducted. Through different places, the situation becomes more realistic as users get to know various stages and a variety of job interview scenarios. Moreover, different rooms for interview scenarios can have entirely different effects on users. Thus they can be used strategically to influence users' interview experience. For example, in an easy version of the interview game, users are welcomed at the reception and then guided into the meeting room, whereas in harder levels, users could initially be seated at the waiting area to raise stress level as they are waiting to be guided into the office of the CEO of the company.

F. Game Fiction Employing Intrinsic Fantasy

Unexpected and unusual concepts have proven to be able to increase engagement of users since they can trigger their curiosity and fantasy. Malone [19] distinguishes between two types of fantasies: intrinsic and extrinsic. In the case of extrinsic fantasy, a problem, e.g., solving a mathematical equation, may be simply overlaid with a game, for example, winning a sports competition. Whether or not gamers make progress toward the goal of the fantasy depends on their abilities to solve the posed problem, but not on events in the fantasy. In the case of intrinsic fantasy, a problem, e.g., learning social skills, is presented as a component of the fantasy world, e.g., interacting with a virtual job interviewer in a three-dimensional (3-D) world. Malone states that intrinsic fantasies are more interesting and more instructional than extrinsic fantasies. In TARDIS and EmpaT, we rely on intrinsic fantasy. That is, there is a close connection between the application of skills and the fantasy world.

A related concept discussed in the literature is curiosity. According to Malone, games can evoke the curiosity by putting users in the environment with "optimal level of information complexity." The environment should be neither too complicated nor too simple concerning the users' existing knowledge. Moreover, it should be novel and surprising, but not incomprehensible. In EmpaT, we increase the user's curiosity by providing them with some initial information on the job but having them discover by themselves details of the job interview (such as the style, format, length, and questions).

G. Immersion and Emotional Involvement

The phenomenon of immersion has been intensely studied in the context of computer games. Immersion roughly relates to the degree of involvement in a game. Bedwell *et al.* [11] link immersion to four attributes that may influence learning progress: objects and agents, representation, sensory stimuli, and safety.

First, the degree of immersion experienced is determined by the objects and agents included in the game scenario. In TARDIS, we did not pay much attention to the environment of the job interview, but only placed the agents into an office room. EmpaT goes beyond TARDIS by including a virtual building of a company that is inhabited by a variety of agents with different roles.

To increase the user's immersion, the agents in the game need to come across as believable. While, for decades, research has concentrated on geometric body modeling and the development of animation and rendering techniques for virtual agents, other qualities have now come in focus as well, including the simulation of conversational and socioemotional behaviors including peculiarities induced by individual personality traits [20]. In order to get immersed in a game, users need to invest emotional energy into the game. Strong emotional involvement may be achieved by a compelling performance of the agents in the game.

In comparison to TARDIS, EmpaT employs nonplayer agents (NPCs) with autonomous behavior and very limited interaction abilities to create a believable background atmosphere (see Fig. 4). For example, on a busy office day, employees meet more frequently. Hence, there is more traffic in the corridor. Furthermore, NPCs can react friendly or harshly when the user passes by adding, even more, possibilities to influence users' emotions (such as anger, frustration, or joy) during the simulated job interview.

Second, the user's sense of immersion depends on representation, i.e., on how realistic the user perceives the gaming environment. To address the aspect of representation, we incorporated findings of organizational and industrial psychology regarding professional job interview procedures, format, and structure. For example, we included common question types, such as situational questions (e.g., "Imagine your department is working with an outdated administration software. By experience, you know a newer alternative. However, your coworkers are critical about this new software. What would you do in this situation?" [21]).

Third, the user's sense of immersion is influenced by sensory stimuli that users perceive during the game experience. We added, among other things, bird sounds, changing lighting conditions throughout the interview process reflecting a

Fig. 4. Locations of the virtual 3-D environment.

changing daytime, and virtual agents walking around talking to each other (see the previously paragraph). These sensory stimuli let users immerse more deeply into the virtual environment as the environment is vivid and changing instead of an entirely sterile environment without any noise.

Fourth, the aspect of safety is defined as a lack of fear toward any negative consequences outside of of the training situation, thus leading to more immersion because users can allow themselves to dive into the situation and try out different strategies without real-world penalties [11]. Indeed, within the game environment, challenging situations might occur in which users feel stress or ashamed, but this experience only enhances the realism of the simulation as these emotions come close to real job interview situations.

In conclusion, we map real-world job interview procedures into a safe virtual environment. This lessens the interview anxiety, elicits emotions in realistic scenarios, and enhances training transfer into real-world job interview situation.

H. Rules/Goals

Rules/goals are defined rules after which to play and objectives that users have to try to achieve within the game [11], [15]. The primary goal within the two job interview scenarios is to complete job interviews successfully using adequate paraverbal and nonverbal behavior. Alongside this goal, the user is confronted with smaller goals throughout the interview, e.g., focus on eye contact during the introduction of the organization or presenting oneself at the beginning of the interview while speaking loud enough and with energetic speech modulation. All these small goals lead the way to the primary aim of succeeding in the complete simulated job interview and eventually to succeed in real-life job interviews. Thus, they motivate and guide users toward improving themselves in applying paraverbal and nonverbal behavior as well as in enhancing declarative and procedural knowledge about job interviews.

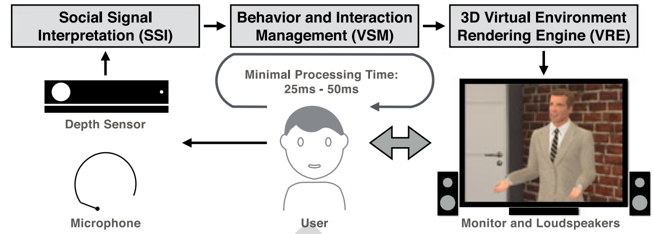

Fig. 5. EmpaT architecture and processing flow.

IV. Architecture

The EmpaT architecture extends the TARDIS architecture by the following several aspects:

1) three-dimensional virtual environment rendering engine instead of 3-D agent rendering engine;
2) extended remote control and logging mechanisms;
3) higher resolution depth camera sensors.

Fig. 5 shows the following main components and the data flow of the architecture:

1) a real-time social signal interpretation framework (SSI);
2) a behavior and interaction modeling and execution tool (VSM) that can be controlled remotely;
3) a 3-D virtual environment rendering engine (VRE) that are asynchronously coordinated with events exchanged by a UDP network architecture.

Each component comes with its own UDP sender and receiver interface. The components SSI, VSM are freely available for research purposes. The VRE component is based on the Unity3D[2] rendering engine, which is also freely available.

The system continually captures, analyzes, and logs the user's voice, gestures, and posture. The minimal processing time for generating a reaction of the current virtual interaction partner is between 25 and 50 ms. The variation in time depends on the amount of signal data of the various communicative channels (voice, gesture, and posture) that have to be analyzed during a user's input action (see Section IV-A). The reaction generation is always triggered by a user's voice action. The generation of nonverbal feedback of the virtual interaction partner (e.g., smiling and nodding backchanneling) starts immediately concerning the above-mentioned timing. The generation of verbal reactions (e.g., comments to a user's input) starts as soon as the user has finished speaking plus a configurable offset of 2 s, in which the user can carry on talking, letting the system wait again. We identified by rule of thumb and by user feedback that 2 s seem to be experienced as an adequate “waiting time.” Future versions of the interaction management will be based on a sophisticated turn-taking model that considers various turn related signals (e.g., gaze and head movement).

The system runs on a high-performance Windows 10 PC with an Intel i7 Hexa-Core at 3.5 GHz, 16 MB Main Memory, and a 2-GB SSD for fast data recording. It requires a high-quality graphics card (NVIDIA GTX 980) and a monitor that is big enough to display the agent in a realistic size (32″). To cancel the environmental noise, the user's voice is recorded with a head

[2] http://unity3d.com

microphone (Sure SM10 and TASKCAM US144-MKII USB Microphone Interface). The Microsoft Kinect II depth sensor captures head movements, gestures, and posture.

A. Social Signal Interpretation

For capturing the user's social cues, we make use of the *Social Signal Interpretation* framework (SSI)[3] [22]. SSI is implemented in C/C++ and makes use of multiple CPU cores. The SSI framework offers tools to record, analyze, and recognize the human behavior, such as gestures, facial expressions, head nods, and emotional speech. Following, a patch-based design pipelines are set up from autonomic components and allow the parallel and synchronized processing of sensor data from multiple input devices. Furthermore, SSI supports machine learning pipelines including the fusion of multiple channels and the synchronization between multiple computers.

For TARDIS and EmpaT, we implemented pipelines that include the detection of the following behavioral cues.

1) Body and facial features: Postures, gestures, head gaze, smiles, motion energy, overall activation.
2) Audio features: Voice activity, intensity, loudness, pitch, audio energy, duration, pulses, periods, unvoiced frames, voice breaks, jitter, shimmer, harmonicity, speech rate.

Besides enabling the system to react to the user in real time, these cues also give us a glimpse into the user's state of mind during the interview, allowing us to observe the impact of the virtual agent's actions on the user.

To compute the audio features intensity, loudness, pitch, and energy we use OpenSMILE [23]. Other features are calculated using algorithms provided by PRAAT [24], [25]. Both systems have been integrated into the SSI Framework to process all features in real time. Relevant parts (e.g., only when the user is speaking) are segmented by voice activity detection to calculate features on utterances of speech. Furthermore, we integrated the Microsoft Speech Platform to our system to allow keyword detection for simple answers and backchanneling, as well as agent and scene control.

B. Behavior and Interaction Management

The behavior and interaction management, the dialog flow, and the content in our application are modeled using the authoring tool *VisualSceneMaker* (VSM) [26]. VSM is programmed in Java and designed precisely to tackle the main challenges that arise when modeling interpersonal coordination [27] and grounding [28] in applications in which social agents interact with humans in situated dialogs and collaborative joint actions.[4] On one hand, it involves the creation of well-aligned multimodal behavior which integrates context knowledge and can automatically be varied in order to avoid repetitive behaviors. On the other hand, it requires the evaluation of temporal and semantic fusion constraints for the incremental recognition of various bidirectional and multimodal behavior patterns. Finally, a fundamental challenge is also the proper coordination, prioritization, and synchronization of a multitude of concurrent, nested, reciprocal, and intertwined processes that are used to implement various behavioral functions on different behavioral levels.

To meet these requirements, the modeling approach with VSM divides the entire modeling process into three largely independent tasks. The authors primarily rely on the following visual and declarative modeling formalisms and textual scripting languages.

1) A textual template-based specification language (comparable to TV and theatre scene scripts) is used for the hybrid creation of knowledge-based and scripted multimodal behavior and dialog content and behavioral activities [29].
2) A logic fact base and logic constraints are used for multimodal fusion and knowledge reasoning as well as asynchronous interprocess communication [30].
3) The dialog and behavior flow, as well as interaction logic, are modeled with a hierarchical and concurrent state-chart variant [31].

Typically, states and transitions are augmented with queries to the logic fact base, playback commands for behavioral activities, and dialog utterances.

The modeling approach of VSM significantly facilitates the distributed and iterative development of clearly structured, easily maintainable and reusable computational dialog, behavior, and interaction models of social agents. The execution environment of VSM pursues an interpreter approach such that its IDE enables an all-time modification and visualization of these models.

C. Interactive 3-D Environment With Virtual agents

Fig. 4 shows a collage of several locations of the EmpaT virtual 3-D environment (VRE) rendered by an extended version of the Unity3D framework.[5]

The virtual environment features the realistic looking 3-D virtual social agents Tom, Tommy, and Susanne[6] (see Fig. 3) besides standard Unity3D virtual agents. They are capable of performing social cue-based interaction with the user. Their lip-sync speech output is using the state-of-the-art Nuance Text-To-Speech system. For a more advanced animation control, they allow the direct manipulation of skeleton model joints (e.g., the neck joint or the spine joint). Also, clothing, hairstyle, accessories, and skin color are customizable. About their communication style, they come with 36 conversational motion-captured gestures (in standing and sitting position), which can be modified during run-time in some aspects (e.g., overall speed, extension, etc.). Besides that, the social agents come with a catalog of 14 facial expressions, which contains among others the six basic emotion expression defined by Ekman [32].

D. Remote Control and Automatic Behavior Annotation

In order to realize a flexible usage of the EmpaT system, all components of the EmpaT system can be remotely controlled (e.g., started, stopped, variable assignment, and message

[3]http://openssi.net
[4]http://scenemaker.dfki.de/
[5]http://www.tricat.net
[6]http://www.charamel.com

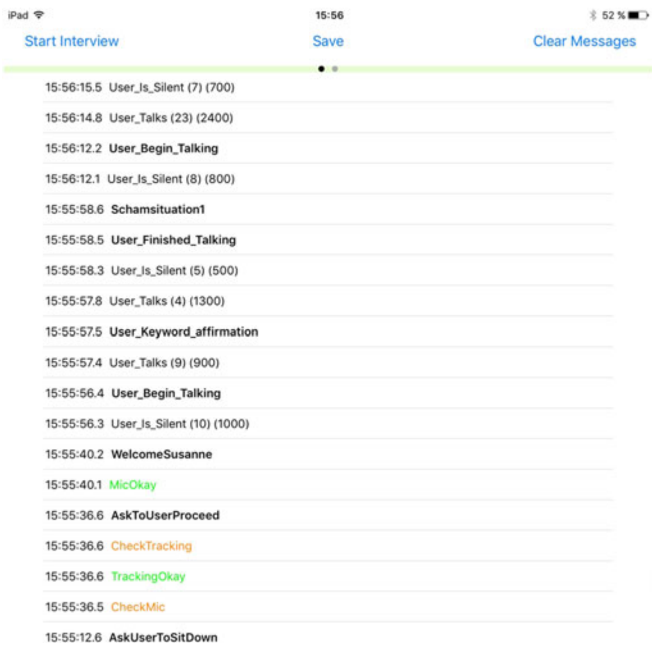

Fig. 6. StudyMaster displays component messages of on-going interactions.

sending). This is realized by the VSM that provides remote control interfaces for all components and to the remote control tool StudyMaster (see Fig. 6).

The StudyMaster exists in two versions: first, an iOS version for Apple iOS devices, written in the Swift programming language,[7] and second, a Java version that runs on every operating system that fully supports Java. StudyMaster receives and sends component messages via a UPD network interface and displays them in a time aligned list. The tool enables to start, to alter, and to observe ongoing interactions. For example:

1) `AskUserToSitDown`—VSM reports that the Scene is performed in which the user is asked to sit down;
2) `TrackingOkay`—SSI reports Kinect tracking is working;
3) `User_Talks`—SSI has detected a user voice signal for more than 200 ms;
4) `User_Is_Silent`—SSI reports the absence of a user's voice signal after the user has talked for a while.

The introduction of a dedicated remote control tool allows study experimenters to fully control the EmpaT game environment without being present in the same room as the participant.

For a postinteraction analysis, we implemented *NovA* [33] (non)verbal Annotator.[8] NovA enables the learners to inspect previous interactions and provides them with an objective report of the social interactions. Typically, different kinds of behaviors are coded on different parallel tracks so that their temporal relationships are clearly visible. Fig. 7 illustrates how NovA determines the level of engagement of an interviewee based on recognized events. In Fig. 7(a), the participant has an open body posture while looking toward the interlocutor and orientating his body in the same direction. In Fig. 7(b), nothing specific

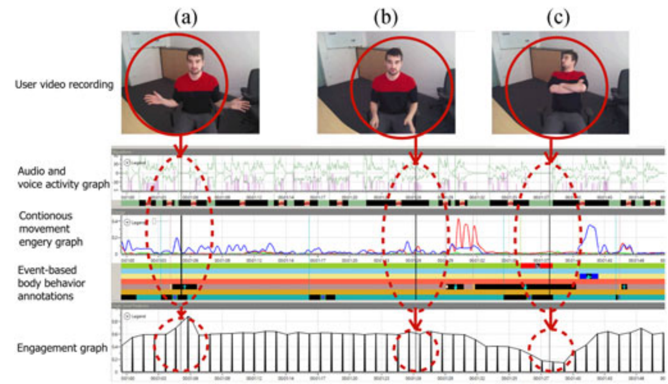

Fig. 7. Comparison of detected cues for (a) high, (b) medium, and (c) low engagement.

is detected, and Fig. 7(c) demonstrates the outcome when the participant uses body language regarded as an indicator of a low amount of engagement, such as leaning back, looking away, and crossing the arms. Bar charts are representing the outcome of the user state recognition for each calculation, which is performed every second.

V. Studies

In the following section, we are going to outline four user studies. The first user study was conducted within the TARDIS project and focused on the core question of a serious game environment: "How does a serious game perform in comparison to traditional learning methods?" In second and third studies, we focused on the agents (TARDIS study) and objects (EmpaT study) influencing the player's emotional reactions. A fourth study (EmpaT study) is about how the virtual environment may influence the users' emotional reaction.

Within the TARDIS project, we conducted an *in situ* study [34] at a local school in Bavaria to investigate the impact of a job interview training game on 20 underprivileged youngsters (10 female) in the age range of 13 to 16. The study was embedded in the existing job interview training of the school. Following a three-day user study, we found that pupils who worked with the training system improved more than those who used traditional learning methods, i.e., reading a written job interview guide. More precisely, professional practitioners rated the overall performance of the pupils who trained with the system significantly better than of those who did not. The system also left a good impression on the school teachers who stated that "using the system, pupils seem to be highly motivated and able to learn how to improve their behavior [...] they usually lack such motivation during class." As a possible reason for this, they mentioned the technical nature of the system, which "transports the experience into the youngster's world" and that the technology-enhanced debriefing phase "makes the feedback much more believable." Pupils also seemed to enjoy interacting with the system. Most of them asked questions regarding how the score was computed, and which of their behaviors contributed to the final score. This suggests that the scoring functionality had a positive effect on the pupils' engagement in the exercise. Furthermore, the game

[7]https://swift.org
[8]http://openssi.net/nova

Fig. 8. Real-time feedback through signal lights (highlighted area shows the magnified version of each signal light).

Fig. 9. Understanding (top) and demanding (bottom) virtual job recruiters.

cards were also received well. One participant even asked for permission to photograph the game cards so she would be able to study them at home.

A second study carried out in the frame of the EmpaT project builds upon the findings of the first study, but it adds some important changes compared to the first study. Most importantly, the game cards are replaced by virtual real time feedback through signal lights on the right side of the screen [2]. These signal lights (see Fig. 8) provided participants with feedback on seven aspects of their nonverbal behavior (smiling, eye contact, posture, arms crossed, nodding, voice volume, and voice energy). In the case of participants expressing adequate nonverbal behavior, the signal light turned green; it turned red if the participants' behavior was not appropriate. It is important to mention that feedback thresholds were based on the psychological literature on nonverbal behavior in general and on nonverbal behavior in interviews.

For example, the threshold for voice volume was 57 dB, which is slightly louder than voice volume in a normal conversation [35]. For other nonverbal behavior, we defined ranges of adequate behavior, for instance in the introduction phase, one to three smiles were defined as adequate, since too less and too much smiling can be detrimental for interview ratings [36] (for detailed information about the definition of the nonverbal feedback, please refer to [2]. During this study, 70 participants (50 female) with a mean age of 24 years from two German universities took part in an interview training study. Participants either received conventional job interview training (i.e., information, pictures, and videos on how to behave during job interviews) or they took part in one round of the EmpaT game; training in both conditions took about 20 min, and participants fulfilled the training on their own and without any support of the experimenter. The crucial difference between the training approaches was that during the EmpaT game, participants actively experienced the interview process in the interaction with the virtual interviewer, and received real-time feedback for their nonverbal behavior using the aforementioned signal lights. After the training, participants answered the measurement of anxiety in selection interviews [37], and then they were interviewed by a trained interviewer. The interviewer assessed participants nonverbal behavior and interview performance in a 20-min semistructured interview. Results showed that participants in the EmpaT game group reported less interview anxiety [$t(68) = 1.67$, $p < 0.05$], they were evaluated as showing more adequate nonverbal behavior [$t(68) = 1.69$, $p < 0.05$], and they received higher interview ratings [$t(68) = 2.50$, $p < 0.05$]; for detailed results consult [2].

A third study that was conducted in the TARDIS project focused on the question of how to increase the level of difficulty by modifying the behavior of the agents in a way that is correlated to the expected level of stress [26]. To this end, we created two profiles of a female virtual job recruiter, understanding, and demanding (see Fig. 9). The former one is defined by letting the agent show narrow gestures close to the body and facial expressions that can be related to positive emotions (e.g., joy, admiration, and happy-for), as well as a friendly head and gaze behavior. Additionally, this agent is using shorter pauses (in comparison to the demanding agent). On the verbal level, explanations and questions show appreciation for the user and contain many politeness phrases. The latter one shows more space-taking (dominant) gestures and facial expressions that can be related to negative emotions (e.g., distress, anger, or reproach), uses longer pauses to show dominance in explanations and questions, and has a dominant gaze behavior.

On the verbal level, comments and questions are strict and contain very few politeness phrases. In the evaluation, 24 participants (7 female) with an average age of 29 years were randomly confronted with the two virtual job recruiters in a simulated job interview. The data included both, subjective measurements in questionnaires and objective measurements like breathing pauses and movement energy. The results of the questionnaires showed that the personality profiles of the virtual agents had an impact on the perceived user experience: the demanding agent induced a higher level of stress than the understanding agent. Participants also felt less comfortable when interacting with the demanding agent and perceived the interview with this agent as more challenging. Furthermore, they rated their performance lower when interacting with this agent. The objective data supported the findings in the questionnaire. The authors interpreted less breathing pauses in the speech and higher movement energy during the demanding condition as a sign for an increased stress level.

Overall, the study shows that it is possible to convey a different learning atmosphere by confronting learners with two opposed agent personalities.

While the third study focused on the impact of the agents on the user's emotional reaction, a fourth study conducted in the EmpaT project investigated how the virtual environment may influence the player's emotional reaction. In TARDIS, the virtual environment consisted only of one room, the room where the interview took place. There was no environment like a company building that could evoke a high degree of immersion in the whole situation. The EmpaT 3-D environment (see Section IV-C) allows us to have participants experience the whole interview situation including the following parts: reaching the company, entering the lobby, announcing one's arrival at the reception, waiting in the reception area, going to the interview room, the actual job interview, and the leaving of the company. During all those steps, participants are confronted with social situations and perceive an atmosphere that has been created with specific research questions in mind. For example, it is possible to manipulate the wall colors and light conditions to find out whether the design of the virtual environment affects the user. This is done in an ongoing study in the EmpaT project. The study tries to give insights about the design of the virtual environment in which a job interview training should take place. We conduct virtual job interviews in the following three different rooms:

1) a neutral one with a neutral wall color and light;
2) an unpleasant one with a dark red wall color and evening light (see Fig. 10, right-hand side);
3) a pleasant one with a friendly light green wall color and bright light like on a sunny day (see Fig. 10, left-hand side).

Measurements include the selection procedural justice scale (SPJS) [38], a measure very commonly used for investigating acceptance of a personnel selection situation (like a job interview), where participants have to assess, for instance, the perceived level of interpersonal treatment and opportunity to perform during the selection interview. Results of the SPJS will indicate, how users experienced the interview itself but also the virtual interviewer. For instance, we hypothesize that an unpleasant room could also reflect the virtual interviewer, who might be perceived less favorable but also to users' perceptions of their performance during the interview. Therefore, participants also have to evaluate their performance, their affective state (emotions, mood), and the virtual room itself.

Fig. 10. Different wall colors and brightness.

These data are not yet entirely available, however, preliminary results show that though the room design does not influence participants' perceptions of the room consciously, the room design seems to affect the assessment of the recruiter as well as the job interview and the self-rated performance. Further analysis of the data will show if the additional evaluation of users' interview performance by a human resource specialist confirms the subjective data, which would point toward a strong influence of the environment on users' behavior.

VI. Conclusion and Future Work

In this paper, we presented an overview of serious game concepts for the design of our serious games. Also, we described the central components of a software platform for creating and researching serious games that support social coaching in the context of job interviews. The platform integrates state-of-the-art technologies for social signal analysis, interaction modeling, and multimodal behavior synthesis. It furthermore incorporates elements from serious game concepts to motivate players and thus increases their willingness to engage in learning.

We presented studies that revealed the benefits of games over books in the context of job interviews. Within two further experiments, we focused on the impact of the agents and the environment on the learner's experience. Within TARDIS, we showed that adaptations of the agents' behavior might induce different levels of stress in the player. Within EmpaT, we demonstrated that even minor changes in the environment, such as changing the room's wall color, may have a measurable effect on the user's learning experience. The two studies revealed that designers of learning environments should be aware that even seeming insignificant attributes might have a significant impact on the learner.

However, a considerable amount of work is still required to further explore the relationship between agents, the virtual environment, and the learner's experience.

ACKNOWLEDGMENT

The authors would like to thank Charamel GmbH and TriCAT GmbH for realizing the requirements with regard to the virtual agents and the 3-D environment.

REFERENCES

[1] K. Anderson *et al.*, "The TARDIS framework: Intelligent virtual agents for social coaching in job interviews," in *Adv. Comput. Entertainment*, D. Reidsma, H. Katayose, and A. Nijholt, Eds. Boekelo, The Netherlands: Springer-Verlag, Nov. 2013, pp. 476–491.
[2] M. Langer, C. J. König, P. Gebhard, and E. André, "Dear computer, teach me manners: Testing virtual employment interview training," *Int. J. Sel. Assessment*, vol. 24, no. 4, pp. 312–323, 2016.
[3] R. Aylett, M. Vala, P. Sequeira, and A. Paiva, "FearNot! – an emergent narrative approach to virtual dramas for anti-bullying education," in *Proc. 4th Int. Conf. Virtual Storytelling. Using Virtual Real. Technol. Storytelling*, Dec. 2007, pp. 202–205.
[4] M. Sapouna *et al.*, "Virtual learning intervention to reduce bullying victimization in primary school: A controlled trial," *J. Child Psychol. Psychiatry*, vol. 51, no. 1, pp. 104–112, 2009.
[5] R. Aylett *et al.*, "Werewolves, cheats, and cultural sensitivity," in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, May 2014, pp. 1085–1092.
[6] S. Thiagarajan and B. Steinwachs, *Barnga: A Simulation Game on Cultural Clashes*, Intercultural Press, 1990.
Q4 [7] B. W. Schuller *et al.*, "Recent developments and results of ASC-Inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions," in *Proc. 3rd Int. Workshop Intell. Digit. Games Empowerment Inclusion*, Mar. 2015.
Q5 [8] X. Pan, M. Gillies, Barker, D. M. C. M. Clark, and M. Slater, "Socially anxious and confident men interact with a forward virtual woman: An experiment study," *PLoS ONE*, vol. 7, no. 4, 2012, Art. no. e32931.
[9] L. M. Batrinca, G. Stratou, A. Shapiro, L. Morency, and S. Scherer, "Cicero - towards a multimodal virtual audience platform for public speaking training," in *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds., vol. 8108. Berlin, Germany: Springer-Verlag, Aug. 2013, pp. 116–128.
[10] M. E. Hoque, M. Courgeon, J. Martin, B. Mutlu, and R. W. Picard, "MACH: My automated conversation coacH," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2013, pp. 697–706.
[11] W. L. Bedwell, D. Pavlas, K. Heyne, E. H. Lazzara, and E. Salas, "Toward a taxonomy linking game attributes to learning an empirical study," *Simul. Gaming*, vol. 43, no. 6, pp. 729–760, 2012.
[12] K. A. Wilson *et al.*, "Relationships between game attributes and learning outcomes: Review and research proposals," *Simul. Gaming*, vol. 40, no. 2, pp. 217–266, May 2008.
[13] M. Mehta, S. Dow, M. Mateas, and B. MacIntyre , "Evaluating a conversation-centered interactive drama," in *Proc. 6th Int. Joint Conf. Auton. Agents Multiagent Systems*, May 2007, Paper 8.
[14] D. Michael and S. Chen, "Proof of learning: Assessment in serious games," 2005. [Online]. Available: https://www.gamasutra.com/view/feature/130843/proof_of_learning_assessment_in_.php
[15] R. Garris, R. Ahlers, and J. E. Driskell, "Games, motivation, and learning: A research and practice model," *Simul. Gaming*, vol. 33, no. 4, pp. 441–467, Dec. 2002.
[16] H. Geser, "Die kommunikative mehrebenenstruktur elementarer interaktionen," *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, vol. 26, pp. 207–231, 1990.
[17] P. Gebhard, T. Baur, I. Damian, G. U. Mehlmann, J. Wagner, and E. André, "Exploring interaction strategies for virtual characters to induce stress in simulated job interviews," in *Proc. Int. Conf. Autonom. Agents Multi-Agent Syst.*, May 2014, pp. 661–668.
[18] B. Endrass, C. Klimmt, G. Mehlmann, E. André, and C. Roth, "Designing user-character dialog in interactive narratives: An exploratory experiment," *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 166–173, Jun. 2014.
[19] T. W. Malone, "What makes things fun to learn? heuristics for designing instructional computer games," in *Proc. 3rd ACM SIGSMALL Symp. 1st SIGPC Symp. Small Syst.*, 1980, pp. 162–169.
[20] E. André and C. Pelachaud, *Interacting With Embodied Conversational Agents*. Boston, MA, USA: Springer-Verlag, 2010, pp. 123–149.
[21] G. P. Latham, L. M. Saari, E. D. Pursell, and M. A. Campion, "The situational interview," *J. Appl. Psychol.*, vol. 65, no. 4, pp. 422–427, 1980.
[22] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, "The social signal interpretation (SSI) framework: Multimodal signal processing and recognition in real-time," in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 831–834.
[23] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the munich open-source multimedia feature extractor," in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 835–838.
[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 4.3.14)," 2005. [Online]. Available: http://www.fon.hum.uva.nl/praat/
[25] N. H. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behav. Res. Methods*, vol. 41, no. 2, pp. 385–390, 2009.
[26] P. Gebhard, G. U. Mehlmann, and M. Kipp, "Visual SceneMaker: A tool for authoring interactive virtual characters," *J. Multimodal User Interfaces, Interact. Embodied Convers. Agents*, vol. 6, no. 1/2, pp. 3–11, 2012.
[27] F. J. Bernieri and R. Rosenthal, *Interpersonal Coordination: Behavior Matching and Interactional Synchrony*. Cambridge, NY, USA: Cambridge Univ. Press, 1991, pp. 401–432.
[28] H. H. Clark, "Coordinating with each other in a material world," *Discourse Stud.*, vol. 7, no. 4/5, pp. 507–525, Oct. 2005.
[29] G. U. Mehlmann, B. Endrass, and E. André, "Modeling parallel state charts for multithreaded multimodal dialogues," in *Proc. 13th Int. Conf. Multimodal Interact.*, 2011, pp. 385–392.
[30] G. U. Mehlmann and E. André, "Modeling multimodal integration with event logic charts," in *Proc. 14th ACM Int. Conf. Multimodal Interact.*, 2012, pp. 125–132.
[31] D. Harel, "Statecharts: A visual formalism for complex systems," *Sci. Comput. Program.*, vol. 8, no. 3, pp. 231–274, Jun. 1987.
[32] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.
[33] T. Baur *et al.*, "Context-aware automated analysis and annotation of social human-agent interactions," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 2, 2015, Art. no. 11.
[34] I. Damian, T. Baur, B. Lugrin, P. Gebhard, G. Mehlmann, and E. André, "Games are better than books: In-situ comparison of an interactive job interview game with conventional training," in *Artificial Intelligence in Education (Lecture Notes in Computer Science)*, vol. 9112. New York, NY, USA: Springer-Verlag, Jun. 2015, pp. 84–94.
[35] H.-P. Zenner, *Die Kommunikation des Menschen: Hören und Sprechen*. Berlin, Germany: Springer, 2011, pp. 334–356.
[36] M. A. Ruben, J. A. Hall, and M. Schmid Mast , "Smiling in a job interview: When less is more," *J. Social Psychol.*, vol. 155, no. 2, pp. 107–126, 2015.
[37] J. McCarthy and R. Goffin, "Measuring job interview anxiety: Beyond weak knees and sweaty palms," *Pers. Psychol.*, vol. 57, no. 3, pp. 607–637, 2004.
[38] T. N. Bauer, D. M. Truxillo, R. J. Sanchez, J. M. Craig, P. Ferrara, and M. A. Campion, "Applicant reactions to selection: Development of the selection procedural justice scale (SPJS)," *Pers. Psychol.*, vol. 54, no. 2, pp. 387–419, 2001.

Patrick Gebhard is the Head of the Affective Computing Group at the German Research Centre for Artificial Intelligence (DFKI), Kaiserslautern, Germany. He has long-term experience in the evaluation, representation, simulation, and display of emotions. His research started over a decade ago, simulating emotions for the creation of believable agent behavior in interactive human-agent systems. He uses this expertise to research human interaction with the goal to enable a humanlike interaction with computer systems. Q6

Tanja Schneeberger received the M.S. degree in psychology from the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany, in 2013.

She is a Researcher in the Affective Computing Group, German Research Centre for Artificial Intelligence, Kaiserslautern, Germany. She is conducting research on user emotions and related nonverbal behavior in dyadic interactions between humans and social agents.

Elisabeth André is a Full Professor in Computer Science at Augsburg University, Augsburg, Germany, and the Chair of the Laboratory for Human-Centered Multimedia. She holds a long track record in embodied conversational agents, multimodal interfaces, and Q7 social signal processing.

Prof. André was elected as a member of the German Academy of Sciences Leopoldina, the Academy of Europe and AcademiaNet. She is also a Fellow of the European Coordinating Committee for Artificial Intelligence.

Tobias Baur is a Researcher at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. His research interests include social signal processing, human-agent interactions, and automated analysis of human nonverbal Q8 behaviors.

Ionut Damian is a Researcher at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. Prior to his graduation in 2011, he conducted research into autonomous agents and virtual environments. His research interests include wearable technology with a focus on signal processing, automatic analysis of human behavior, and Q9 intelligent feedback design.

Gregor Mehlmann received the B.S. degree and the Master of Computer Science honors degree from Saarland University, Saarbrücken, Germany. He is currently working toward the Ph.D. degree. Q10

He is a Researcher and Lecturer at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. Prior to this, he was a Research Associate at the German Research Centre for Artificial Intelligence.

Cornelius König is a Full Professor at the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany. One of his research interests is the use of latest computer science developments for training and personnel selection. Q11

Mr. König is the President of the section for Work, Organizational, and Business Psychology within the German Psychological Society.

Markus Langer is a Researcher at the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany. In his research, he is connecting computer science and psychology. Specifically, he is conducting research on nonverbal behavior and novel technologies for human resource management processes like personnel selection and training, for example, virtual characters as interviewers, and recognition of and automatic feedback for nonverbal behavior during job interview training. Q12

QUERIES

Q1. Author: Please verify that the funding information is correct.pdf.

Q2. Author: Please supply index terms/keywords for your paper. To download the IEEE Taxonomy go to https://www.ieee.org/documents/taxonomy_v101.pdf.

Q3. Author: Please check the edits made in the sentence "We added, among...talking to each other."

Q4. Author: Please provide the location of the publisher in Ref. [6]. Also check whether the reference is ok as set.

Q5. Author: Please provide the page range for Ref. [7].

Q6. Author: Please provide the full educational details (degree, subject, etc.) of P. Gebhard.

Q7. Author: Please provide the full educational details (degree, subject, etc.) of E. André.

Q8. Author: Please provide the full educational details (degree, subject, etc.) of T. Baur.

Q9. Author: Please provide the full educational details (degree, subject, etc.) of I. Damian.

Q10. Author: Please provide the subject, year, and educational institute (name/location) in which G. Mehlmann received the Bachelor's degree, the subject and year in which he received the Master of Computer Science degree, and the subject and educational institute details in which he is currently working toward the Ph.D. degree.

Q11. Author: Please provide the full educational details (degree, subject, etc.) of C. König.

Q12. Author: Please provide the full educational details (degree, subject, etc.) of M. Langer.

Serious Games for Training Social Skills in Job Interviews

Patrick Gebhard, Tanja Schneeberger, Elisabeth André, Tobias Baur, Ionut Damian, Gregor Mehlmann, Cornelius König, and Markus Langer

***Abstract*—In this paper, we focus on experience-based role play with virtual agents to provide young adults at the risk of exclusion with social skill training. We present a scenario-based serious game simulation platform. It comes with a social signal interpretation component, a scripted and autonomous agent dialog and social interaction behavior model, and an engine for 3-D rendering of lifelike virtual social agents in a virtual environment. We show how two training systems developed on the basis of this simulation platform can be used to educate people in showing appropriate socioemotive reactions in job interviews. Furthermore, we give an overview of four conducted studies investigating the effect of the agents' portrayed personality and the appearance of the environment on the players' perception of the characters and the learning experience.**

***Index Terms*—.**

I. Introduction

PEDAGOGICAL role play with virtual agents offers great promise for social skill training. It provides learners with a realistic, but safe environment that enables them to train specific verbal and nonverbal behaviors in order to adapt to socially challenging situations. At the same time, learners benefit from the gamelike environment, which increases not only their enjoyment and motivation but also enables them to take a step back from the environment and think about their behavior if necessary.

In this paper, we will present a scenario-based serious game simulation platform that supports social training and coaching in the context of job interviews. The game simulation platform has been developed in the TARDIS project [1] and further extended in the EmpaT project [2]. The platform includes technology to detect the users' emotions and social attitudes in real time through voice, gestures, and facial expressions during the interaction with a virtual agent as a job interviewer. To achieve their pedagogical goals, TARDIS and EmpaT need to expose the players to situations in the learning environment that evoke similar reactions in them as real job interviews. They require a high demand for computational intelligence and perceptual skills in order to understand the player's socioemotional reactions and optimally adapt the pace of learning.

Fig. 1. User interacting with TARDIS. Paperboard cards give hints on how to behave for each phase of a job interview.

In TARDIS, users were able to interact with a virtual recruiter that responded to their paraverbal and nonverbal behaviors (see Fig. 1). However, users were not immersed in the physical setting in which the job interview took place (e.g., the building and the room style, the employees, or the specific atmospheric setup). Furthermore, the TARDIS users' experience is limited to the job interview setup, in which the user sits in front of the virtual job recruiter at a desk.

EmpaT embeds the job interview into a virtual environment that comes with a virtual personal assistant who explains every step of the job interview experience. Moreover, the virtual environment allows simulating various challenges that come along with job interviews, as that users may navigate through to find the room where the actual job interview will take place (see Fig. 2). On their way to the interview, users arrive to the reception desk asking for the job interview appointment and wait until they are called for the interview in the nearby lobby. In

Manuscript received December 30, 2016; revised August 31, 2017; accepted January 13, 2018. Date of publication; date of current version. This work was supported in part by the German Ministry of Education and Research (BMBF) within the EmpaT project (funding code 16SV7229K) and in part by the European Commission within FP7-ICT-2011-7 (project TARDIS, Grant Agreement 288578). *(Corresponding author: Patrick Gebhard.)*

P. Gebhard and T. Schneeberger are with the German Research Centre for Artificial Intelligence, Saarbrücken 66123, Germany (e-mail: gebhard@dfki.de; tanja.schneeberger@dfki.de).

E. André, T. Baur, I. Damian, and G. Mehlmann are with the Augsburg University, Augsburg 86159, Germany (e-mail: andre@informatik.uni-augsburg.de; baur@hcm-lab.de; damian@hcm-lab.de; gregor.mehlmann@informatik.uni-augsburg.de).

C. König and M. Langer are with the Saarland University, Saarbrücken 66123, Germany (e-mail: ckoenig@mx.uni-saarland.de; markus.langer@uni-saarland.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TG.2018.2808525

Fig. 2. Company building in EmpaT, in which the interview takes place.

the waiting phase, users can observe the daily routine of the simulated employees. The EmpaT system allows confronting users with situations that might increase their uneasiness, for example, when having to ask unfriendly personnel for directions or in case of interruptions during the actual job interview. Thus, EmpaT enables a more comprehensive experience that includes all phases of a job interview from entering to leaving the building of the company where the job interview takes place.

In the following, we will first discuss related work on the use of computer-enhanced role play for social coaching. After that, we will analyze elements of game design that may have an impact on the achievement of pedagogical goals in social coaching. We then present the serious game simulation platform that supports social learning in the context of job interviews. Finally, we present four studies we conducted to investigate the impact of serious games for social skill training and the influence of the agents' behaviors and the physical environment on the players' perception of the agents and the learning experience.

II. Related Work

Computerized social skill training tools have seen rapid evolution in recent years due to advances in the areas of social signal processing as well as improvements in the audio-visual rendering of virtual agents. Such tools are meant to complement or even substitute traditional training approaches.

A variety of serious games employs role play with virtual agents that foster reflection about socioemotional interactions. An example includes the anti-bullying Game FearNot! that has been developed within the project eCircus [3]. The project investigates how social learning may be enhanced through interactive role play with virtual agents that establish empathetic relationships with the learners. It creates interactive stories in a virtual school with embodied conversational agents in the role of bullies, helpers, victims, etc. The children run through various bullying episodes, interact with the virtual agents after each episode, and provide advice to them. The benefit of educational role plays of this kind lies in the fact that they promote reflective thinking. Results of a conducted evaluation [4] showed that the system had a positive effect on the children's abilities to cope with bullying.

Role play with virtual agents has also been a popular approach to educate users in cultural sensitivity. Employing role play with virtual agents that represent different cultures, users are supposed to develop a better understanding of other cultures. Eventually, the users are expected to develop intercultural empathy and reduce their negative attitude toward other cultures. An example of such a system has been developed within the eCute project: The objective of MIXER (moderating interactions for cross-cultural empathic relationships)[1] is to enable users to experience emotions that are usually elicited during interactions of members of a different group [5]. To this end, children are confronted with scenarios in which virtual agents appear to violate previously introduced rules in a game scenario. Such a situation leads inevitably to frustration and negative attitudes toward members of the other group. By interacting with MIXED, children are expected to learn to reflect about behaviors of other groups and reconsider potentially existing prejudices against them. The setting was inspired by the card-game BARNGA, which has been successfully used for cultural training of adults [6]. Other than expected by the authors, the MIXER game did not foster cultural awareness in children in a pilot study. The authors assumed that the learning objectives MIXER was designed to meet were not appropriate for the age group that was not able to cope with the negative rule-clash-based conflict.

While the above-described systems analyze the user's verbal and nonverbal behaviors for the purpose of the interaction, their primary objective is to help users cope with socially challenging situations. They do not aim at teaching users appropriate socioemotional communication skills directly.

Within the project ASD-Inclusion [7], techniques for the recognition of human socioemotional behaviors have been employed to help children who have autism to improve their socioemotional communication skills. A game platform with virtual agents has been developed that enables children to learn how emotions can be expressed and recognized via gestures, facial, and vocal expressions in a virtual game world. A requirement analysis revealed the need to incorporate an appropriate incentive system to keep children engaged. Therefore, the authors implemented a monetary system which rewarded children with virtual money for improved performance from which they could buy items for their avatars.

Furthermore, social signal processing techniques have been employed to automatically record and analyze the learner's social and emotional signals, whereas virtual agents are employed to simulate various social situations, such as social gatherings [8] or public speeches [9]. Similar to our work is a job interview simulation with a virtual agent by Hoque *et al.* [10]. They explored the impact of the job interview training environment on MIT students and concluded that students who used the system to train, experienced a larger performance increase than students who used conventional methods. These results are encouraging for our research. However, while Hoque *et al.* recruited MIT students as participants, our target group are job-seeking young

[1] http://ecute.eu/mixer/

people who have been categorized as being at risk of exclusion. Furthermore, they did not explicitly incorporate elements from games to increase the players' motivation.

A number of studies reveal the positive effects of gamelike environments for social coaching. However, the research conducted in the eCute project also points out difficulties in designing a gamelike environment that achieves particular pedagogical goals. Overall, there is still a lack of knowledge on the relationship between specific game attributes and learning outcomes. In the next section, we use the taxonomy by Bedwell and colleagues [11] as a starting point for the analysis of game attributes in TARDIS and EmpaT.

III. Game Experience

To support social coaching in TARDIS and EmpaT, we incorporated elements from serious games for which we hypothesized a positive effect on learning. To this end, we consulted the work by Wilson *et al.* [12] as well as Bedwell *et al.* [11] who identified eight categories of game attributes designers should be especially aware of when developing gamified environments: action language, assessment, conflict/challenge, control, environment, game fiction, human interaction, immersion, and rules/goals. In the remainder of this section, we take a closer look upon seven of these game attribute categories (we will not include human interaction, as there is no human interaction in the two job interview training games) and describe to what extent they have been taken into account during the design of the job interview games in TARDIS and EmpaT.

A. Nonverbal and Paraverbal Behavior as an "Action Language"

Action language defines the way how users interact with the game (e.g., by using a joystick or a keyboard). It is an important aspect to consider when designing gamified environments as the mode of interaction may have a strong influence on the learning outcome [12]. In commercial computer games, the action language employed to communicate with the game represents a well-defined mapping between commands to be input by the user and actions to be executed by the game. Unlike commercial games, TARDIS and EmpaT rely on natural forms of interaction with focus on paraverbal and nonverbal behavior to which the interview agents react in a believable manner.

This form of interaction poses particular challenges to the design of the interaction. Due to deficiencies of current technology to process natural language input, effective strategies had to be found to support a consistent and coherent conversational flow. Based on an evaluation of Façade, a gamelike interactive storytelling scenario with conversational agents, Mehta *et al.* [13] came up with a number of guidelines and recommendations for dialogue design in gamelike environments, such as avoiding shallow confirmations of user input and supporting the user's abilities to make sense of recognition flaws. Both in TARDIS and in EmpaT, the user is supposed to play a role that is in accordance by the learning goals. To support a smooth conversational flow, the virtual agents provide explicit interaction prompts. That is the agents are modeled in a way that they are requesting specific information from the user. This way, the user knows what kind of input is required and learns at the same time which questions are typically asked in a job interview. As long as the user follows the rules of the game, there is no need to conduct a deep semantic analysis of the user's utterances even though some simple form of keyword spotting has shown beneficial. Due to the design of the scenario, failures of the natural language understanding technologies could be interpreted as communication issues that typically arise in job interviews. For example, a virtual job interviewer shifting to another topic due to natural language understanding problems may still provide a compelling performance, for example, by indicating boredom of the previous topic. Text-based input would facilitate the analysis of natural language input significantly. However, this option had to be discarded in our case. First, text-based input would break the illusion of a realistic job interview. Second, users are expected to acquire appropriate paraverbal and nonverbal behaviors that have to be synchronized with their speech. Consequently, the game environment should enable them to practice these behaviors.

B. Assessment Through Social Sensing

Assessment refers to the feedback given to the user on their progress [14]. In order to keep users motivated, it is essential to provide feedback to them on how well they are doing so far and how advanced they are regarding specific goals [11]. In a social setting with virtual agents, direct feedback can be given naturally by the agents' nonverbal and verbal cues. However, users might not always understand such implicit cues. Learning to read somebody's body language could be the topic of a serious game on its own, but would distract from the actual learning goals here. In order to increase the agents' believability in TARDIS and EmpaT, they respond immediately to the user's input by appropriate nonverbal and verbal cues. However, we also incorporated more explicit feedback in TARDIS and EmpaT that helps users improve their behavior in subsequent interactions.

In TARDIS, we implemented a reward system that remunerates users after execution of successful actions. To encourage adequate behaviors, the system scores the users' performance and rewards him or her with points if he or she behaves in compliance with behaviors specified on a game card (see Fig. 1). A score for the user's behavior is computed in real time during the interaction by using sensing devices to recognize social cues, such as a smile or crossed arms. Providing feedback on social behavior is an ambitious task due to the high amount of subjectivity and lack of transparency. For example, it may be counterproductive to tell the user that he or she appears disengaged without giving him or her the reasons for such an assessment. Therefore, TARDIS offers additional feedback to users in a debriefing phase through a graphical user interface that highlights social cues that contributed to the system's assessment of the user's behavior (see Section IV-D).

In EmpaT, we are currently exploring possibilities of giving users continuous feedback on their behavior and progress. The challenge consists in providing such feedback without disturbing the flow of the game. Currently, we are investigating the use of signal lights to give feedback on paraverbal and nonverbal behavior dynamically and in real time. For example, the signal light for eye contact would turn red if someone is not keeping eye contact with the interviewer for a predefined ratio of time, but the signal light would adapt dynamically and turn green again if the participant succeeds in keeping eye contact for longer than the above-mentioned ratio of time. Furthermore, we are studying immediate reactions of the virtual interview agent to the users' behavior, such as exhorting users if they interrupt the virtual agent during its speech. This kind of assessment raises awareness of how to behave during job interviews and enable them to learn how to apply nonverbal behavior adequately. Furthermore, positive feedback improves the users' self-efficacy and enhances their motivation to keep on training social skills behaviors.

C. *Different Levels of Conflict/Challenge*

Adding conflict/challenge leads to difficulties and problems within the game that need to be solved, as well as to uncertainties enhancing the tension. For instance, random events like employees coming into the interview room and disturbing the interaction can add unforeseeable aspects. Another example would be that participants can be confronted with job interview questions of varying difficulty enhancing replayability. Thus, conflict/challenge is a driving force within the game that keeps the users motivated to proceed [11], [15]. It is important to note that it is crucial to define difficulty levels carefully, so the game is sufficiently challenging, but not too difficult [12].

Within TARDIS and EmpaT, we implemented various levels of difficulty offering a challenging experience for users with different levels of job interview experience.

TARDIS makes use of one virtual agent with different social behavior profiles, understanding and demanding, which consequently influence the level of difficulty of the simulation as well as the impact on the user.

In EmpaT, job interviews are performed by one out of two virtual interviewers of different age: a young and middle-aged male, and a 50-years old female (see Fig. 3, center and right-hand sides) reflecting experience and status of the agent [16]. Furthermore, these agents express different nonverbal and verbal behaviors which portray the agents' personality (understanding, demanding, and neutral) [17]. Depending on their personality profile, these agents evoke emotions in the user that are experienced in real job interviews and thus enhance the realism of the simulation (see Section V). Also, the EmpaT realization provides users with an understanding personal assistant that guides the user through the interview experience (see Fig. 3, left-hand side).

In addition to increasing the level of difficulty by agents representing a higher status, EmpaT introduces critical events in the job interview. For instance, in an entry level job interview, there is a young interview agent in casual clothing behaving in amiable manner and asking easy and common interview questions. In comparison, at a higher level, the age and appearance of the interview agent reflect a more experienced member of the organization or even the leader of the company. Questions in the higher level job interview are less common or even provoking. Thus, interviewees have to adapt to the enhanced degree of difficulty through different behavior. Also, random events can be added. For example, another virtual agent might enter the room or the interviewer might make a challenging comment on the user's behavior. This way, the game can be modulated to create tension and stress in the users similarly to a real job interview situation, thus enhancing the realism of the simulation. Providing challenges to the users can lead to reduced anxiety in real job interview situations and improved self-efficacy because the users already have experienced similar situations in the training game. Moreover, customizable difficulty and random events enhance replayability, further increasing exposure to the training environment.

Fig. 3. Virtual 3-D environment (VRE) social agents.

D. *Guided Control*

Control describes how much users can influence the game by their actions [11], [15]. A high level of control can positively impact the users' experience, but it can also be detrimental if users get lost within the environment [11]. Within the EmpaT job interview training, the user can walk around freely to explore the virtual environment. However, at some point, the user will be led to the meeting room by the virtual interviewer.

When designing the dialog with the virtual interviewer, the question arises of how much control should be given to the user. A mixed-initiative dialog gives more freedom to the user. However, it also requires more sophisticated language understanding capabilities than system-initiative dialog. In [18], we compared the system-initiative dialog with mixed-initiative dialog in a soap-opera-like game environment that included a text input interface to enable users to communicate with virtual agents. The users preferred the mixed-initiative dialog over the system-initiative dialogue even though the mixed-initiative dialog was less robust. Apparently, the experiential advantages of the mixed-initiative dialog compensated for the lower amount of accuracy in natural language understanding.

TARDIS and EmpaT rely on a speech-based input which comes with even greater challenges than a text-based input. Therefore, we decided to implement the less demanding option of system-initiative dialog in order to ensure a smooth flow of dialogue. This interaction style appears to match the situation of a job interview well where the applicants are not expected to take over control. Furthermore, the system-initiative dialog still gives autonomy to the users. During the actual interview, users can focus on the main aspects of the simulation: the questions the interviewer asks, their answers, and their paraverbal and nonverbal behavior—still leaving a high level of control to users through speech and body movement. Thus, the simulation and its outcomes depend on users' own actions. This setup enhances realism and gives users the opportunity to experiment with their nonverbal behavior and learn about consequences.

E. Realistic Environment

The environment defines where users find themselves in the game and how they see the world [11]. In EmpaT, users see the world in first person view as they walk through a realistic office building. The entrance hall of the company building has a reception desk, where users are welcomed by a virtual agent, a waiting room where users wait to be picked up by the interview agent, and various rooms where the interview can be conducted. Through different places, the situation becomes more realistic as users get to know various stages and a variety of job interview scenarios. Moreover, different rooms for interview scenarios can have entirely different effects on users. Thus they can be used strategically to influence users' interview experience. For example, in an easy version of the interview game, users are welcomed at the reception and then guided into the meeting room, whereas in harder levels, users could initially be seated at the waiting area to raise stress level as they are waiting to be guided into the office of the CEO of the company.

F. Game Fiction Employing Intrinsic Fantasy

Unexpected and unusual concepts have proven to be able to increase engagement of users since they can trigger their curiosity and fantasy. Malone [19] distinguishes between two types of fantasies: intrinsic and extrinsic. In the case of extrinsic fantasy, a problem, e.g., solving a mathematical equation, may be simply overlaid with a game, for example, winning a sports competition. Whether or not gamers make progress toward the goal of the fantasy depends on their abilities to solve the posed problem, but not on events in the fantasy. In the case of intrinsic fantasy, a problem, e.g., learning social skills, is presented as a component of the fantasy world, e.g., interacting with a virtual job interviewer in a three-dimensional (3-D) world. Malone states that intrinsic fantasies are more interesting and more instructional than extrinsic fantasies. In TARDIS and EmpaT, we rely on intrinsic fantasy. That is, there is a close connection between the application of skills and the fantasy world.

A related concept discussed in the literature is curiosity. According to Malone, games can evoke the curiosity by putting users in the environment with "optimal level of information complexity." The environment should be neither too complicated nor too simple concerning the users' existing knowledge. Moreover, it should be novel and surprising, but not incomprehensible. In EmpaT, we increase the user's curiosity by providing them with some initial information on the job but having them discover by themselves details of the job interview (such as the style, format, length, and questions).

G. Immersion and Emotional Involvement

The phenomenon of immersion has been intensely studied in the context of computer games. Immersion roughly relates to the degree of involvement in a game. Bedwell *et al.* [11] link immersion to four attributes that may influence learning progress: objects and agents, representation, sensory stimuli, and safety.

First, the degree of immersion experienced is determined by the objects and agents included in the game scenario. In TARDIS, we did not pay much attention to the environment of the job interview, but only placed the agents into an office room. EmpaT goes beyond TARDIS by including a virtual building of a company that is inhabited by a variety of agents with different roles.

To increase the user's immersion, the agents in the game need to come across as believable. While, for decades, research has concentrated on geometric body modeling and the development of animation and rendering techniques for virtual agents, other qualities have now come in focus as well, including the simulation of conversational and socioemotional behaviors including peculiarities induced by individual personality traits [20]. In order to get immersed in a game, users need to invest emotional energy into the game. Strong emotional involvement may be achieved by a compelling performance of the agents in the game.

In comparison to TARDIS, EmpaT employs nonplayer agents (NPCs) with autonomous behavior and very limited interaction abilities to create a believable background atmosphere (see Fig. 4). For example, on a busy office day, employees meet more frequently. Hence, there is more traffic in the corridor. Furthermore, NPCs can react friendly or harshly when the user passes by adding, even more, possibilities to influence users' emotions (such as anger, frustration, or joy) during the simulated job interview.

Second, the user's sense of immersion depends on representation, i.e., on how realistic the user perceives the gaming environment. To address the aspect of representation, we incorporated findings of organizational and industrial psychology regarding professional job interview procedures, format, and structure. For example, we included common question types, such as situational questions (e.g., "Imagine your department is working with an outdated administration software. By experience, you know a newer alternative. However, your coworkers are critical about this new software. What would you do in this situation?" [21]).

Third, the user's sense of immersion is influenced by sensory stimuli that users perceive during the game experience. We added, among other things, bird sounds, changing lighting conditions throughout the interview process reflecting a

Fig. 4. Locations of the virtual 3-D environment.

changing daytime, and virtual agents walking around talking to each other (see the previously paragraph). These sensory stimuli let users immerse more deeply into the virtual environment as the environment is vivid and changing instead of an entirely sterile environment without any noise.

Fourth, the aspect of safety is defined as a lack of fear toward any negative consequences outside of the training situation, thus leading to more immersion because users can allow themselves to dive into the situation and try out different strategies without real-world penalties [11]. Indeed, within the game environment, challenging situations might occur in which users feel stress or ashamed, but this experience only enhances the realism of the simulation as these emotions come close to real job interview situations.

In conclusion, we map real-world job interview procedures into a safe virtual environment. This lessens the interview anxiety, elicits emotions in realistic scenarios, and enhances training transfer into real-world job interview situation.

H. Rules/Goals

Rules/goals are defined rules after which to play and objectives that users have to try to achieve within the game [11], [15]. The primary goal within the two job interview scenarios is to complete job interviews successfully using adequate paraverbal and nonverbal behavior. Alongside this goal, the user is confronted with smaller goals throughout the interview, e.g., focus on eye contact during the introduction of the organization or presenting oneself at the beginning of the interview while speaking loud enough and with energetic speech modulation. All these small goals lead the way to the primary aim of succeeding in the complete simulated job interview and eventually to succeed in real-life job interviews. Thus, they motivate and guide users toward improving themselves in applying paraverbal and nonverbal behavior as well as in enhancing declarative and procedural knowledge about job interviews.

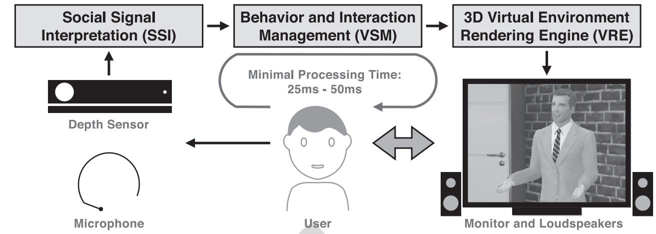

Fig. 5. EmpaT architecture and processing flow.

IV. Architecture

The EmpaT architecture extends the TARDIS architecture by the following several aspects:

1) three-dimensional virtual environment rendering engine instead of 3-D agent rendering engine;
2) extended remote control and logging mechanisms;
3) higher resolution depth camera sensors.

Fig. 5 shows the following main components and the data flow of the architecture:

1) a real-time social signal interpretation framework (SSI);
2) a behavior and interaction modeling and execution tool (VSM) that can be controlled remotely;
3) a 3-D virtual environment rendering engine (VRE) that are asynchronously coordinated with events exchanged by a UDP network architecture.

Each component comes with its own UDP sender and receiver interface. The components SSI, VSM are freely available for research purposes. The VRE component is based on the Unity3D[2] rendering engine, which is also freely available.

The system continually captures, analyzes, and logs the user's voice, gestures, and posture. The minimal processing time for generating a reaction of the current virtual interaction partner is between 25 and 50 ms. The variation in time depends on the amount of signal data of the various communicative channels (voice, gesture, and posture) that have to be analyzed during a user's input action (see Section IV-A). The reaction generation is always triggered by a user's voice action. The generation of nonverbal feedback of the virtual interaction partner (e.g., smiling and nodding backchanneling) starts immediately concerning the above-mentioned timing. The generation of verbal reactions (e.g., comments to a user's input) starts as soon as the user has finished speaking plus a configurable offset of 2 s, in which the user can carry on talking, letting the system wait again. We identified by rule of thumb and by user feedback that 2 s seem to be experienced as an adequate "waiting time." Future versions of the interaction management will be based on a sophisticated turn-taking model that considers various turn related signals (e.g., gaze and head movement).

The system runs on a high-performance Windows 10 PC with an Intel i7 Hexa-Core at 3.5 GHz, 16 MB Main Memory, and a 2-GB SSD for fast data recording. It requires a high-quality graphics card (NVIDIA GTX 980) and a monitor that is big enough to display the agent in a realistic size (32′′). To cancel the environmental noise, the user's voice is recorded with a head

[2]http://unity3d.com

microphone (Sure SM10 and TASKCAM US144-MKII USB Microphone Interface). The Microsoft Kinect II depth sensor captures head movements, gestures, and posture.

A. Social Signal Interpretation

For capturing the user's social cues, we make use of the *Social Signal Interpretation* framework (SSI)[3] [22]. SSI is implemented in C/C++ and makes use of multiple CPU cores. The SSI framework offers tools to record, analyze, and recognize the human behavior, such as gestures, facial expressions, head nods, and emotional speech. Following, a patch-based design pipelines are set up from autonomic components and allow the parallel and synchronized processing of sensor data from multiple input devices. Furthermore, SSI supports machine learning pipelines including the fusion of multiple channels and the synchronization between multiple computers.

For TARDIS and EmpaT, we implemented pipelines that include the detection of the following behavioral cues.

1) Body and facial features: Postures, gestures, head gaze, smiles, motion energy, overall activation.
2) Audio features: Voice activity, intensity, loudness, pitch, audio energy, duration, pulses, periods, unvoiced frames, voice breaks, jitter, shimmer, harmonicity, speech rate.

Besides enabling the system to react to the user in real time, these cues also give us a glimpse into the user's state of mind during the interview, allowing us to observe the impact of the virtual agent's actions on the user.

To compute the audio features intensity, loudness, pitch, and energy we use OpenSMILE [23]. Other features are calculated using algorithms provided by PRAAT [24], [25]. Both systems have been integrated into the SSI Framework to process all features in real time. Relevant parts (e.g., only when the user is speaking) are segmented by voice activity detection to calculate features on utterances of speech. Furthermore, we integrated the Microsoft Speech Platform to our system to allow keyword detection for simple answers and backchanneling, as well as agent and scene control.

B. Behavior and Interaction Management

The behavior and interaction management, the dialog flow, and the content in our application are modeled using the authoring tool *VisualSceneMaker* (VSM) [26]. VSM is programmed in Java and designed precisely to tackle the main challenges that arise when modeling interpersonal coordination [27] and grounding [28] in applications in which social agents interact with humans in situated dialogs and collaborative joint actions.[4] On one hand, it involves the creation of well-aligned multimodal behavior which integrates context knowledge and can automatically be varied in order to avoid repetitive behaviors. On the other hand, it requires the evaluation of temporal and semantic fusion constraints for the incremental recognition of various bidirectional and multimodal behavior patterns. Finally, a fundamental challenge is also the proper coordination, prioritization, and synchronization of a multitude of concurrent, nested, reciprocal, and intertwined processes that are used to implement various behavioral functions on different behavioral levels.

To meet these requirements, the modeling approach with VSM divides the entire modeling process into three largely independent tasks. The authors primarily rely on the following visual and declarative modeling formalisms and textual scripting languages.

1) A textual template-based specification language (comparable to TV and theatre scene scripts) is used for the hybrid creation of knowledge-based and scripted multimodal behavior and dialog content and behavioral activities [29].
2) A logic fact base and logic constraints are used for multimodal fusion and knowledge reasoning as well as asynchronous interprocess communication [30].
3) The dialog and behavior flow, as well as interaction logic, are modeled with a hierarchical and concurrent state-chart variant [31].

Typically, states and transitions are augmented with queries to the logic fact base, playback commands for behavioral activities, and dialog utterances.

The modeling approach of VSM significantly facilitates the distributed and iterative development of clearly structured, easily maintainable and reusable computational dialog, behavior, and interaction models of social agents. The execution environment of VSM pursues an interpreter approach such that its IDE enables an all-time modification and visualization of these models.

C. Interactive 3-D Environment With Virtual agents

Fig. 4 shows a collage of several locations of the EmpaT virtual 3-D environment (VRE) rendered by an extended version of the Unity3D framework.[5]

The virtual environment features the realistic looking 3-D virtual social agents Tom, Tommy, and Susanne[6] (see Fig. 3) besides standard Unity3D virtual agents. They are capable of performing social cue-based interaction with the user. Their lip-sync speech output is using the state-of-the-art Nuance Text-To-Speech system. For a more advanced animation control, they allow the direct manipulation of skeleton model joints (e.g., the neck joint or the spine joint). Also, clothing, hairstyle, accessories, and skin color are customizable. About their communication style, they come with 36 conversational motion-captured gestures (in standing and sitting position), which can be modified during run-time in some aspects (e.g., overall speed, extension, etc.). Besides that, the social agents come with a catalog of 14 facial expressions, which contains among others the six basic emotion expression defined by Ekman [32].

D. Remote Control and Automatic Behavior Annotation

In order to realize a flexible usage of the EmpaT system, all components of the EmpaT system can be remotely controlled (e.g., started, stopped, variable assignment, and message

[3]http://openssi.net
[4]http://scenemaker.dfki.de/
[5]http://www.tricat.net
[6]http://www.charamel.com

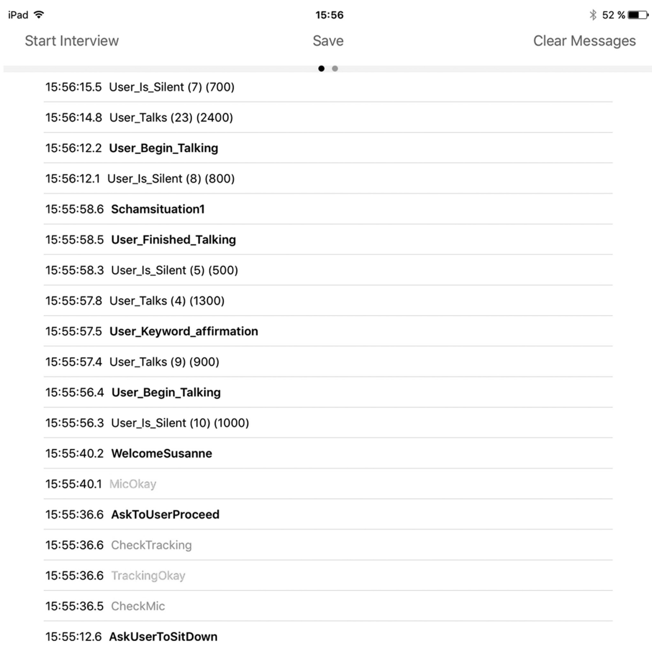

Fig. 6. StudyMaster displays component messages of on-going interactions.

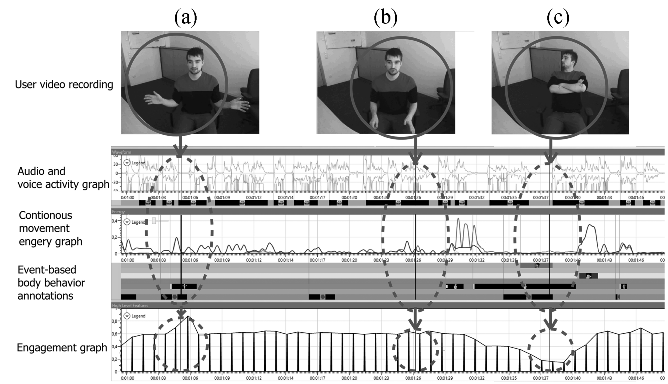

Fig. 7. Comparison of detected cues for (a) high, (b) medium, and (c) low engagement.

sending). This is realized by the VSM that provides remote control interfaces for all components and to the remote control tool StudyMaster (see Fig. 6).

The StudyMaster exists in two versions: first, an iOS version for Apple iOS devices, written in the Swift programming language,[7] and second, a Java version that runs on every operating system that fully supports Java. StudyMaster receives and sends component messages via a UPD network interface and displays them in a time aligned list. The tool enables to start, to alter, and to observe ongoing interactions. For example:

1) `AskUserToSitDown`—VSM reports that the Scene is performed in which the user is asked to sit down;
2) `TrackingOkay`—SSI reports Kinect tracking is working;
3) `User_Talks`—SSI has detected a user voice signal for more than 200 ms;
4) `User_Is_Silent`—SSI reports the absence of a user's voice signal after the user has talked for a while.

The introduction of a dedicated remote control tool allows study experimenters to fully control the EmpaT game environment without being present in the same room as the participant.

For a postinteraction analysis, we implemented *NovA* [33] (non)verbal Annotator.[8] NovA enables the learners to inspect previous interactions and provides them with an objective report of the social interactions. Typically, different kinds of behaviors are coded on different parallel tracks so that their temporal relationships are clearly visible. Fig. 7 illustrates how NovA determines the level of engagement of an interviewee based on recognized events. In Fig. 7(a), the participant has an open body posture while looking toward the interlocutor and orientating his body in the same direction. In Fig. 7(b), nothing specific is detected, and Fig. 7(c) demonstrates the outcome when the participant uses body language regarded as an indicator of a low amount of engagement, such as leaning back, looking away, and crossing the arms. Bar charts are representing the outcome of the user state recognition for each calculation, which is performed every second.

[7]https://swift.org
[8]http://openssi.net/nova

V. Studies

In the following section, we are going to outline four user studies. The first user study was conducted within the TARDIS project and focused on the core question of a serious game environment: "How does a serious game perform in comparison to traditional learning methods?" In second and third studies, we focused on the agents (TARDIS study) and objects (EmpaT study) influencing the player's emotional reactions. A fourth study (EmpaT study) is about how the virtual environment may influence the users' emotional reaction.

Within the TARDIS project, we conducted an *in situ* study [34] at a local school in Bavaria to investigate the impact of a job interview training game on 20 underprivileged youngsters (10 female) in the age range of 13 to 16. The study was embedded in the existing job interview training of the school. Following a three-day user study, we found that pupils who worked with the training system improved more than those who used traditional learning methods, i.e., reading a written job interview guide. More precisely, professional practitioners rated the overall performance of the pupils who trained with the system significantly better than of those who did not. The system also left a good impression on the school teachers who stated that "using the system, pupils seem to be highly motivated and able to learn how to improve their behavior [...] they usually lack such motivation during class." As a possible reason for this, they mentioned the technical nature of the system, which "transports the experience into the youngster's world" and that the technology-enhanced debriefing phase "makes the feedback much more believable." Pupils also seemed to enjoy interacting with the system. Most of them asked questions regarding how the score was computed, and which of their behaviors contributed to the final score. This suggests that the scoring functionality had a positive effect on the pupils' engagement in the exercise. Furthermore, the game

Fig. 8. Real-time feedback through signal lights (highlighted area shows the magnified version of each signal light).

Fig. 9. Understanding (top) and demanding (bottom) virtual job recruiters.

cards were also received well. One participant even asked for permission to photograph the game cards so she would be able to study them at home.

A second study carried out in the frame of the EmpaT project builds upon the findings of the first study, but it adds some important changes compared to the first study. Most importantly, the game cards are replaced by virtual real time feedback through signal lights on the right side of the screen [2]. These signal lights (see Fig. 8) provided participants with feedback on seven aspects of their nonverbal behavior (smiling, eye contact, posture, arms crossed, nodding, voice volume, and voice energy). In the case of participants expressing adequate nonverbal behavior, the signal light turned green; it turned red if the participants' behavior was not appropriate. It is important to mention that feedback thresholds were based on the psychological literature on nonverbal behavior in general and on nonverbal behavior in interviews.

For example, the threshold for voice volume was 57 dB, which is slightly louder than voice volume in a normal conversation [35]. For other nonverbal behavior, we defined ranges of adequate behavior, for instance in the introduction phase, one to three smiles were defined as adequate, since too less and too much smiling can be detrimental for interview ratings [36] (for detailed information about the definition of the nonverbal feedback, please refer to [2]. During this study, 70 participants (50 female) with a mean age of 24 years from two German universities took part in an interview training study. Participants either received conventional job interview training (i.e., information, pictures, and videos on how to behave during job interviews) or they took part in one round of the EmpaT game; training in both conditions took about 20 min, and participants fulfilled the training on their own and without any support of the experimenter. The crucial difference between the training approaches was that during the EmpaT game, participants actively experienced the interview process in the interaction with the virtual interviewer, and received real-time feedback for their nonverbal behavior using the aforementioned signal lights. After the training, participants answered the measurement of anxiety in selection interviews [37], and then they were interviewed by a trained interviewer. The interviewer assessed participants nonverbal behavior and interview performance in a 20-min semistructured interview. Results showed that participants in the EmpaT game group reported less interview anxiety [$t(68) = 1.67$, $p < 0.05$], they were evaluated as showing more adequate nonverbal behavior [$t(68) = 1.69$, $p < 0.05$], and they received higher interview ratings [$t(68) = 2.50$, $p < 0.05$]; for detailed results consult [2].

A third study that was conducted in the TARDIS project focused on the question of how to increase the level of difficulty by modifying the behavior of the agents in a way that is correlated to the expected level of stress [26]. To this end, we created two profiles of a female virtual job recruiter, understanding, and demanding (see Fig. 9). The former one is defined by letting the agent show narrow gestures close to the body and facial expressions that can be related to positive emotions (e.g., joy, admiration, and happy-for), as well as a friendly head and gaze behavior. Additionally, this agent is using shorter pauses (in comparison to the demanding agent). On the verbal level, explanations and questions show appreciation for the user and contain many politeness phrases. The latter one shows more space-taking (dominant) gestures and facial expressions that can be related to negative emotions (e.g., distress, anger, or reproach), uses longer pauses to show dominance in explanations and questions, and has a dominant gaze behavior.

On the verbal level, comments and questions are strict and contain very few politeness phrases. In the evaluation, 24 participants (7 female) with an average age of 29 years were randomly confronted with the two virtual job recruiters in a simulated job interview. The data included both, subjective measurements in questionnaires and objective measurements like breathing pauses and movement energy. The results of the questionnaires showed that the personality profiles of the virtual agents had an impact on the perceived user experience: the demanding agent induced a higher level of stress than the understanding agent. Participants also felt less comfortable when interacting with the demanding agent and perceived the interview with this agent as more challenging. Furthermore, they rated their performance lower when interacting with this agent. The objective data supported the findings in the questionnaire. The authors interpreted less breathing pauses in the speech and higher movement energy during the demanding condition as a sign for an increased stress level.

Overall, the study shows that it is possible to convey a different learning atmosphere by confronting learners with two opposed agent personalities.

While the third study focused on the impact of the agents on the user's emotional reaction, a fourth study conducted in the EmpaT project investigated how the virtual environment may influence the player's emotional reaction. In TARDIS, the virtual environment consisted only of one room, the room where the interview took place. There was no environment like a company building that could evoke a high degree of immersion in the whole situation. The EmpaT 3-D environment (see Section IV-C) allows us to have participants experience the whole interview situation including the following parts: reaching the company, entering the lobby, announcing one's arrival at the reception, waiting in the reception area, going to the interview room, the actual job interview, and the leaving of the company. During all those steps, participants are confronted with social situations and perceive an atmosphere that has been created with specific research questions in mind. For example, it is possible to manipulate the wall colors and light conditions to find out whether the design of the virtual environment affects the user. This is done in an ongoing study in the EmpaT project. The study tries to give insights about the design of the virtual environment in which a job interview training should take place. We conduct virtual job interviews in the following three different rooms:

1) a neutral one with a neutral wall color and light;
2) an unpleasant one with a dark red wall color and evening light (see Fig. 10, right-hand side);
3) a pleasant one with a friendly light green wall color and bright light like on a sunny day (see Fig. 10, left-hand side).

Measurements include the selection procedural justice scale (SPJS) [38], a measure very commonly used for investigating acceptance of a personnel selection situation (like a job interview), where participants have to assess, for instance, the perceived level of interpersonal treatment and opportunity to perform during the selection interview. Results of the SPJS will indicate, how users experienced the interview itself but also the virtual interviewer. For instance, we hypothesize that an unpleasant room could also reflect the virtual interviewer, who might be perceived less favorable but also to users' perceptions of their performance during the interview. Therefore, participants also have to evaluate their performance, their affective state (emotions, mood), and the virtual room itself.

Fig. 10. Different wall colors and brightness.

These data are not yet entirely available, however, preliminary results show that though the room design does not influence participants' perceptions of the room consciously, the room design seems to affect the assessment of the recruiter as well as the job interview and the self-rated performance. Further analysis of the data will show if the additional evaluation of users' interview performance by a human resource specialist confirms the subjective data, which would point toward a strong influence of the environment on users' behavior.

VI. Conclusion and Future Work

In this paper, we presented an overview of serious game concepts for the design of our serious games. Also, we described the central components of a software platform for creating and researching serious games that support social coaching in the context of job interviews. The platform integrates state-of-the-art technologies for social signal analysis, interaction modeling, and multimodal behavior synthesis. It furthermore incorporates elements from serious game concepts to motivate players and thus increases their willingness to engage in learning.

We presented studies that revealed the benefits of games over books in the context of job interviews. Within two further experiments, we focused on the impact of the agents and the environment on the learner's experience. Within TARDIS, we showed that adaptations of the agents' behavior might induce different levels of stress in the player. Within EmpaT, we demonstrated that even minor changes in the environment, such as changing the room's wall color, may have a measurable effect on the user's learning experience. The two studies revealed that designers of learning environments should be aware that even seeming insignificant attributes might have a significant impact on the learner.

However, a considerable amount of work is still required to further explore the relationship between agents, the virtual environment, and the learner's experience.

Acknowledgment

The authors would like to thank Charamel GmbH and TriCAT GmbH for realizing the requirements with regard to the virtual agents and the 3-D environment.

References

[1] K. Anderson *et al.*, "The TARDIS framework: Intelligent virtual agents for social coaching in job interviews," in *Adv. Comput. Entertainment*, D. Reidsma, H. Katayose, and A. Nijholt, Eds. Boekelo, The Netherlands: Springer-Verlag, Nov. 2013, pp. 476–491.

[2] M. Langer, C. J. König, P. Gebhard, and E. André, "Dear computer, teach me manners: Testing virtual employment interview training," *Int. J. Sel. Assessment*, vol. 24, no. 4, pp. 312–323, 2016.

[3] R. Aylett, M. Vala, P. Sequeira, and A. Paiva, "FearNot! – an emergent narrative approach to virtual dramas for anti-bullying education," in *Proc. 4th Int. Conf. Virtual Storytelling. Using Virtual Real. Technol. Storytelling*, Dec. 2007, pp. 202–205.

[4] M. Sapouna *et al.*, "Virtual learning intervention to reduce bullying victimization in primary school: A controlled trial," *J. Child Psychol. Psychiatry*, vol. 51, no. 1, pp. 104–112, 2009.

[5] R. Aylett *et al.*, "Werewolves, cheats, and cultural sensitivity," in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, May 2014, pp. 1085–1092.

[6] S. Thiagarajan and B. Steinwachs, *Barnga: A Simulation Game on Cultural Clashes*, Intercultural Press, 1990.

[7] B. W. Schuller *et al.*, "Recent developments and results of ASC-Inclusion: An integrated internet-based environment for social inclusion of children with autism spectrum conditions," in *Proc. 3rd Int. Workshop Intell. Digit. Games Empowerment Inclusion*, Mar. 2015.

[8] X. Pan, M. Gillies, Barker, D. M. C. M. Clark, and M. Slater, "Socially anxious and confident men interact with a forward virtual woman: An experiment study," *PLoS ONE*, vol. 7, no. 4, 2012, Art. no. e32931.

[9] L. M. Batrinca, G. Stratou, A. Shapiro, L. Morency, and S. Scherer, "Cicero - towards a multimodal virtual audience platform for public speaking training," in *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, Eds., vol. 8108. Berlin, Germany: Springer-Verlag, Aug. 2013, pp. 116–128.

[10] M. E. Hoque, M. Courgeon, J. Martin, B. Mutlu, and R. W. Picard, "MACH: My automated conversation coacH," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2013, pp. 697–706.

[11] W. L. Bedwell, D. Pavlas, K. Heyne, E. H. Lazzara, and E. Salas, "Toward a taxonomy linking game attributes to learning an empirical study," *Simul. Gaming*, vol. 43, no. 6, pp. 729–760, 2012.

[12] K. A. Wilson *et al.*, "Relationships between game attributes and learning outcomes: Review and research proposals," *Simul. Gaming*, vol. 40, no. 2, pp. 217–266, May 2008.

[13] M. Mehta, S. Dow, M. Mateas, and B. MacIntyre , "Evaluating a conversation-centered interactive drama," in *Proc. 6th Int. Joint Conf. Auton. Agents Multiagent Systems*, May 2007, Paper 8.

[14] D. Michael and S. Chen, "Proof of learning: Assessment in serious games," 2005. [Online]. Available: https://www.gamasutra.com/view/feature/130843/proof_of_learning_assessment_in_.php

[15] R. Garris, R. Ahlers, and J. E. Driskell, "Games, motivation, and learning: A research and practice model," *Simul. Gaming*, vol. 33, no. 4, pp. 441–467, Dec. 2002.

[16] H. Geser, "Die kommunikative mehrebenenstruktur elementarer interaktionen," *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, vol. 26, pp. 207–231, 1990.

[17] P. Gebhard, T. Baur, I. Damian, G. U. Mehlmann, J. Wagner, and E. André, "Exploring interaction strategies for virtual characters to induce stress in simulated job interviews," in *Proc. Int. Conf. Autonom. Agents Multi-Agent Syst.*, May 2014, pp. 661–668.

[18] B. Endrass, C. Klimmt, G. Mehlmann, E. André, and C. Roth, "Designing user-character dialog in interactive narratives: An exploratory experiment," *IEEE Trans. Comput. Intell. AI Games*, vol. 6, no. 2, pp. 166–173, Jun. 2014.

[19] T. W. Malone, "What makes things fun to learn? heuristics for designing instructional computer games," in *Proc. 3rd ACM SIGSMALL Symp. 1st SIGPC Symp. Small Syst.*, 1980, pp. 162–169.

[20] E. André and C. Pelachaud, *Interacting With Embodied Conversational Agents*. Boston, MA, USA: Springer-Verlag, 2010, pp. 123–149.

[21] G. P. Latham, L. M. Saari, E. D. Pursell, and M. A. Campion, "The situational interview," *J. Appl. Psychol.*, vol. 65, no. 4, pp. 422–427, 1980.

[22] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, "The social signal interpretation (SSI) framework: Multimodal signal processing and recognition in real-time," in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 831–834.

[23] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in openSMILE, the munich open-source multimedia feature extractor," in *Proc. 21st ACM Int. Conf. Multimedia*, 2013, pp. 835–838.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 4.3.14)," 2005. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[25] N. H. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behav. Res. Methods*, vol. 41, no. 2, pp. 385–390, 2009.

[26] P. Gebhard, G. U. Mehlmann, and M. Kipp, "Visual SceneMaker: A tool for authoring interactive virtual characters," *J. Multimodal User Interfaces, Interact. Embodied Convers. Agents*, vol. 6, no. 1/2, pp. 3–11, 2012.

[27] F. J. Bernieri and R. Rosenthal, *Interpersonal Coordination: Behavior Matching and Interactional Synchrony*. Cambridge, NY, USA: Cambridge Univ. Press, 1991, pp. 401–432.

[28] H. H. Clark, "Coordinating with each other in a material world," *Discourse Stud.*, vol. 7, no. 4/5, pp. 507–525, Oct. 2005.

[29] G. U. Mehlmann, B. Endrass, and E. André, "Modeling parallel state charts for multithreaded multimodal dialogues," in *Proc. 13th Int. Conf. Multimodal Interact.*, 2011, pp. 385–392.

[30] G. U. Mehlmann and E. André, "Modeling multimodal integration with event logic charts," in *Proc. 14th ACM Int. Conf. Multimodal Interact.*, 2012, pp. 125–132.

[31] D. Harel, "Statecharts: A visual formalism for complex systems," *Sci. Comput. Program.*, vol. 8, no. 3, pp. 231–274, Jun. 1987.

[32] P. Ekman, "An argument for basic emotions," *Cogn. Emotion*, vol. 6, no. 3/4, pp. 169–200, 1992.

[33] T. Baur *et al.*, "Context-aware automated analysis and annotation of social human-agent interactions," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 2, 2015, Art. no. 11.

[34] I. Damian, T. Baur, B. Lugrin, P. Gebhard, G. Mehlmann, and E. André, "Games are better than books: In-situ comparison of an interactive job interview game with conventional training," in *Artificial Intelligence in Education (Lecture Notes in Computer Science)*, vol. 9112. New York, NY, USA: Springer-Verlag, Jun. 2015, pp. 84–94.

[35] H.-P. Zenner, *Die Kommunikation des Menschen: Hören und Sprechen*. Berlin, Germany: Springer, 2011, pp. 334–356.

[36] M. A. Ruben, J. A. Hall, and M. Schmid Mast , "Smiling in a job interview: When less is more," *J. Social Psychol.*, vol. 155, no. 2, pp. 107–126, 2015.

[37] J. McCarthy and R. Goffin, "Measuring job interview anxiety: Beyond weak knees and sweaty palms," *Pers. Psychol.*, vol. 57, no. 3, pp. 607–637, 2004.

[38] T. N. Bauer, D. M. Truxillo, R. J. Sanchez, J. M. Craig, P. Ferrara, and M. A. Campion, "Applicant reactions to selection: Development of the selection procedural justice scale (SPJS)," *Pers. Psychol.*, vol. 54, no. 2, pp. 387–419, 2001.

Patrick Gebhard is the Head of the Affective Computing Group at the German Research Centre for Artificial Intelligence (DFKI), Kaiserslautern, Germany. He has long-term experience in the evaluation, representation, simulation, and display of emotions. His research started over a decade ago, simulating emotions for the creation of believable agent behavior in interactive human-agent systems. He uses this expertise to research human interaction with the goal to enable a humanlike interaction with computer systems.

Tanja Schneeberger received the M.S. degree in psychology from the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany, in 2013.

She is a Researcher in the Affective Computing Group, German Research Centre for Artificial Intelligence, Kaiserslautern, Germany. She is conducting research on user emotions and related nonverbal behavior in dyadic interactions between humans and social agents.

Elisabeth André is a Full Professor in Computer Science at Augsburg University, Augsburg, Germany, and the Chair of the Laboratory for Human-Centered Multimedia. She holds a long track record in embodied conversational agents, multimodal interfaces, and social signal processing. Q7

Prof. André was elected as a member of the German Academy of Sciences Leopoldina, the Academy of Europe and AcademiaNet. She is also a Fellow of the European Coordinating Committee for Artificial Intelligence.

Tobias Baur is a Researcher at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. His research interests include social signal processing, human-agent interactions, and automated analysis of human nonverbal behaviors. Q8

Ionut Damian is a Researcher at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. Prior to his graduation in 2011, he conducted research into autonomous agents and virtual environments. His research interests include wearable technology with a focus on signal processing, automatic analysis of human behavior, and intelligent feedback design. Q9

Gregor Mehlmann received the B.S. degree and the Master of Computer Science honors degree from Saarland University, Saarbrücken, Germany. He is currently working toward the Ph.D. degree. Q10

He is a Researcher and Lecturer at the Human Centered Multimedia Laboratory, Augsburg University, Augsburg, Germany. Prior to this, he was a Research Associate at the German Research Centre for Artificial Intelligence.

Cornelius König is a Full Professor at the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany. One of his research interests is the use of latest computer science developments for training and personnel selection. Q11

Mr. König is the President of the section for Work, Organizational, and Business Psychology within the German Psychological Society.

Markus Langer is a Researcher at the Department of Work and Organizational Psychology, Saarland University, Saarbrücken, Germany. In his research, he is connecting computer science and psychology. Specifically, he is conducting research on nonverbal behavior and novel technologies for human resource management processes like personnel selection and training, for example, virtual characters as interviewers, and recognition of and automatic feedback for nonverbal behavior during job interview training. Q12

QUERIES

Q1. Author: Please verify that the funding information is correct.pdf.
Q2. Author: Please supply index terms/keywords for your paper. To download the IEEE Taxonomy go to https://www.ieee.org/documents/taxonomy_v101.pdf.
Q3. Author: Please check the edits made in the sentence "We added, among...talking to each other."
Q4. Author: Please provide the location of the publisher in Ref. [6]. Also check whether the reference is ok as set.
Q5. Author: Please provide the page range for Ref. [7].
Q6. Author: Please provide the full educational details (degree, subject, etc.) of P. Gebhard.
Q7. Author: Please provide the full educational details (degree, subject, etc.) of E. André.
Q8. Author: Please provide the full educational details (degree, subject, etc.) of T. Baur.
Q9. Author: Please provide the full educational details (degree, subject, etc.) of I. Damian.
Q10. Author: Please provide the subject, year, and educational institute (name/location) in which G. Mehlmann received the Bachelor's degree, the subject and year in which he received the Master of Computer Science degree, and the subject and educational institute details in which he is currently working toward the Ph.D. degree.
Q11. Author: Please provide the full educational details (degree, subject, etc.) of C. König.
Q12. Author: Please provide the full educational details (degree, subject, etc.) of M. Langer.