Exploring Interaction Strategies for Virtual Characters to Induce Stress in Simulated Job Interviews

Patrick Gebhard
DFKI GmbH, Campus D3 2
Saarbrücken, Germany
gebhard@dfki.de

Tobias Baur
Human Centered Multimedia
Augsburg University, Germany
baur@hcm-lab.de

Ionut Damian
Human Centered Multimedia
Augsburg University, Germany
damian@hcm-lab.de

Gregor Mehlmann
Human Centered Multimedia
Augsburg University, Germany
mehlmann@hcm-lab.de

Johannes Wagner
Human Centered Multimedia
Augsburg University, Germany
wagner@hcm-lab.de

Elisabeth André
Human Centered Multimedia
Augsburg University, Germany
andre@hcm-lab.de

ABSTRACT

Job interviews come with a number of challenges, especially for young people who are out of employment, education, or training (NEETs). This paper presents an approach to a job training simulation environment that employs two virtual characters and social cue recognition techniques to create an immersive interactive job interview. The two virtual characters are created with different social behavior profiles, understanding and demanding, which consequently influences the level of difficulty of the simulation as well as the impact on the user. In this context we present a user study which investigates the feasibility of the proposed approach by measuring the effect the different virtual characters have on users.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information System**]: Artificial, Augmented, and Virtual Realities

General Terms

Design, Experimentation

Keywords

Social Training, Social Virtual Characters

1. INTRODUCTION

In this paper we present a simulated interactive job interview for young people who are out of employment, education, or training (NEETs). NEETs often have underdeveloped socio-emotional and interaction skills. This comes along with a low self-confidence, a lack of sense of their own strengths, or social anxiety [19]. As a consequence, they may experience problems in challenging situations, such as job interviews, where they need to convince the recruiter of their fit in a company. In order to help NEETS train strategies for stress management, we confront them with virtual characters in the role of recruiters. Previous work has shown that the behavior of a virtual character may have an impact on the user's perceived level of stress. For example, Prendinger and Ishizuka [23] report on a study that revealed that a user's stress may be reduced if a virtual character apologizes for a delay caused by the machine.

Figure 1: Facial expressions and gestures of the understanding and demanding virtual recruiter.

The objective of our work is to adapt the level of interaction difficulties for the user by modifying the behavior of the characters in a way that is correlated to the expected level of stress. Creating characters that induce various levels of stress in the user is a challenge. On the one hand, the char-

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

acter's behavior has to be sufficiently articulated in order to have an impact on the user. On the other hand, extreme behaviors have to be avoided in order to convey a convincing performance. We focus on two character types: 1) an *understanding* and 2) a *demanding* recruiter. The purpose of the *understanding* character is to help job seekers get acquainted to the situation of a job interview. This character aims to enhance the learners' self efficacy by creating a warm and friendly atmosphere during the simulation. In contrast, the *demanding* character serves to prepare the learners for more difficult situations during a job interview and train their coping abilities. More precisely, the character tries to induce stress in the learners by conveying a cold and unfriendly atmosphere. With a study that especially takes the user's social cues during the simulated interview into account, we investigate: 1) how users experience the different characters' personality, 2) differences in induced stress by them, and 3) the impact on the users' interaction behavior.

2. RELATED WORK

In order to help people train social skills, a variety of techniques have been developed, such as role playing, group discussions or specific exercises [15]. The need for effective social training has also inspired a number of proposals for computer-based simulation environments as additional platforms for delivering such training for a variety of applications including job interviews [23, 16], public speeches [2], social situations [22] inter-cultural communication [10], negotiation scenarios [26] or psychotherapy [17]. Some of these approaches confront learners with different kinds of characters, but their verbal and nonverbal behavior has not been adjusted in a systematic manner in order to meet specific learning goals.

A number of attempts have been made to adapt a virtual character's verbal and nonverbal behaviors in such a way that it conveys a particular personality or status. One of the earliest approaches includes the work by André and colleagues [1] who presented a team of virtual presenters that was able to vary linguistic style according to chosen personality traits. While André and colleagues [1] used a simple template-based generator, Mairesse and Walker [20] developed a more sophisticated language generator for conveying different personality traits that included modules for content planning, sentence planning, and surface realization. Kipp and Gebhard [18] investigated different gaze strategies to convey the intended degree of dominance and submission. The submissive character only looked briefly every now and then at the human interlocutor and immediately averted the gaze again. On the contrary to the submissive character, the dominant character established and held eye contact when speaking, and after speaking, immediately looked away. Bee and colleagues [5] investigated the dominance perception of virtual character with varying head and eye gaze directions, and found that a lowered head was perceived less dominant than a raised head. Pollock and colleagues [6] investigated how the combination of gaze and linguistic behaviors impacts the perception of social dominance. They found that the linguistic expression of disagreeableness had a significant effect on dominance perception, but that extraversion did not. This work is complemented by a study conducted by Neff and colleagues [21] who focused on the interactions between natural language and gestures. Based on psychological findings, they showed how a virtual character's gestural and linguistic behaviors can be modulated to adjust the perceived degree of extraversion. Very relevant to the presented work is the recently created corpus of virtual character's non-verbal behaviors collected by users; categorized and prepared for further research by Ravent, Ochs, and Pelachaud [24]. They started by letting users relate non-verbal behavior of a virtual character to its interpersonal attitudes. Then, these data are analyzed with regard to how a social virtual character behaves in interactive situations. One result of their work is a Bayesian network using these data to create a computational model of non-verbal behavior depending on the interpersonal attitude of the character and its actual (social) intentions. For this work, we are not using this network itself, but we rely on several behavioral aspects that we transfer to our social virtual characters.

Our work is inspired by the research question how the behavior of a virtual coach may be adapted in order to adjust the degree of interaction difficulty in staged job interviews. Earlier work in the area of computer-based tutoring systems focused on the design of conversational strategies that lower the learner's stress level. For example, Prendinger and Ishizuka [23] designed a virtual job recruiter that analyzes the users' physiological data in order to provide empathic feedback, such as "*It seems you did not like this question so much!*". In contrast, our work aims to prepare human learners to challenging social interactions by confronting them with interactive virtual characters that elicit a different amount of stress in them.

3. ADAPTING THE BEHAVIOR OF THE TWO VIRTUAL CHARACTERS

The situation of a job interview is likely to elicit stress in the user. This stress may be increased by specifc verbal behaviors of the recruiter, such as asking unpleasant questions or making rude remarks. However, nonverbal behaviors of the recruiter, such as looking skeptically, may cause stress as well. In our work, we exploit multiple channels of expression to elicit different levels of stress in the user.

We designed 2 job recruiters. Each of them exhibit a different personality and related non-verbal and verbal behavior aspects that influence the level of stress, which a user might experience during the interaction:

Understanding. The non-verbal behavior of the understanding character is defined by letting the character 1) show narrow gestures close to the body, 2) show facial expressions that can be related to positive emotions (e.g. joy, admiration, happy-for), 3) using shorter pauses (in comparison to the demanding character), and 4) show a friendly head and gaze behavior. On the verbal level, explanations and questions show appreciation for the user and contain many politeness phrases.

Demanding. The non-verbal behavior of the demanding character is defined as: 1) show more space-taking (dominant) gestures, 2) show facial expressions that can be related to negative emotions (e.g. distress, anger, or reproach), 3) using longer pauses to show dominance in explanations and questions, and 4) show a dominant gaze behavior. On the verbal level, comments and questions are strict and contain very few politeness phrases.

To control the amount of stress by linguistic variation, our work starts from the politeness theory developed by Brown and Levinson [8]. According to the politeness theory, all social actors have *face wants*: the desire for positive *face* (being approved of by others) and the desire for negative *face* (being unimpeded by others). Many conversational exchanges between people, (e.g., offers, requests, commands) potentially threaten positive *face*, negative *face*, or both. To avoid this, speakers employ various types of face threat mitigation strategies to reduce the impact on *face*. Strategies identified by Brown and Levinson include positive politeness (emphasizing approval of the interlocutor), negative politeness (emphasizing the interlocutor's freedom of action, e.g., via a suggestion) and off-record statements (indirect statements that imply that an action is needed).

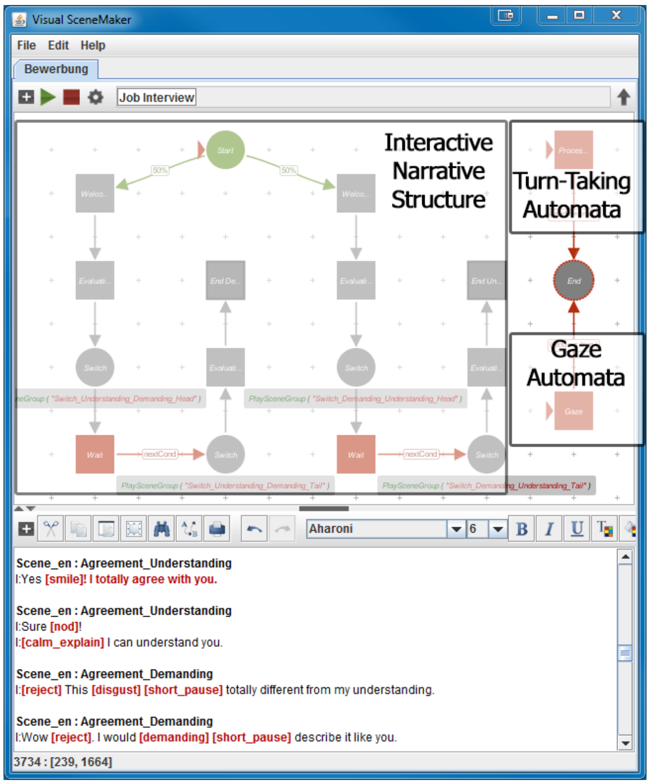

Figure 2: Main dialog plot, Turn-Taking, and Gaze-Behavior modeled as parallel HSFMs.

In job interviews, a job seeker's face may be easily threatened by showing disapproval ("*Your answer does not yet persuade me.*") or by demonstrating power over him or her ("*I give you five minutes to convince me about your skills.*"). Such behaviors are likely to cause stress in the job seeker that might have a negative impact on their performance. By experiencing such a behavior in a simulation, the job seeker may explore various strategies to cope with stressful situations. In our work, we adapt the character's degree of politeness by modifying dialogue acts by the degree of attention paid to the job seeker's face. For example, the demanding character conveys its messages bluntly to the user ("*Your final grades at school seem to be below average.*"), the understanding character applies a variety of strategies to mitigate potential face threats, such as convey interest ("*I am eager to know more about ...*"), claim in-group membership ("*Let us now talk about ...*") or give option not to act ("*Would you like to tell me ...*").

Furthermore, the level of stress is modulated by the character's backchannel behavior. Backchannels are brief verbal or nonverbal cues listeners provide during a conversation to indicate the speaker that they are still following the conversation. Previous work [14] shows that the simulation of backchannel cues helps increase the level of rapport between a virtual character and a human. To create a pleasant atmosphere for the user, the understanding character signals comprehension by head nods and brief verbal utterances, such as "*Ok*" or "*I see*". The user is also encouraged through the use of positive feedback, such as smiles. In contrast, the demanding character does not provide any backchannel cues to indicate engagement in the interaction. Additionally, it tries to unsettle the user by showing negative feedback, such as frowning.

In addition to the character's backchannel behavior, we use different types of conversational gestures and key posture features for realizing the two personalities. We rely on the work of Ravenet, Ochs, and Pelachaud who have collected a corpus of a virtual characters non-verbal behavior that a character should display to convey particular interpersonal attitudes [24]. The corpus contains a comprehensive overview on non-verbal and verbal behavior rules, which are congruent with the related literature we use. For this work, we focus on the rules for dominant and friendly gestures (e.g. dominant gestures are characterized with a large spatial parameter) and posture feature (e.g. tilt of the head on a side with no gaze not averted).

To further increase the user's stress level, the demanding character exhibits a gaze behavior that is supposed to be perceived as dominant, whereas the understanding character shows a less dominant gaze behavior that is supposed to be perceived as unobstrusive. In total 2 gaze strategies are implemented: *dominant* and *friendly*. The dominant strategy follows Kipp and Gebhard's iGaze approach [18]. It consists of maintaining eye contact at the beginning of a speaking or listening phase. First, the character establishes and holds eye contact and after a first utterance, it looks away for a short time. The friendly gaze strategy follows the findings from Fukayama et al. [12].

4. THE SYSTEM

For our interactive scenario we rely on the SEMAINE API [25], a framework for enabling the creation of emotion-oriented systems with virtual characters. The system features a real time social cue recognition system, a scenario manager, and the Charamel 3D character rendering environment[1]. As shown in Figure 1, the main user interface is the Gloria character that is capable of performing social cue-based interaction with the user. She is able to perform lip-sync speech output using the state-of-the-art Nuance Text-To-Speech system. For a more advanced animation control, Gloria allows the direct manipulation of her skeleton model joints (e.g. the neck joint or the spine joint). Gloria comes with 54 conversational motion-captured gestures, which can be modified in during run-time in some aspects (e.g. overall speed, extension ...). In addition, the character comes with a catalog of 14 facial expressions, which contains among others the 6 basic emotion expression defined by Ekman [9].

Our system allows us to record a user's social cues by a Mi-

[1]http://www.charamel.com

crosoft Kinect device and to analyze by the Social Cue Recognizer at run-time. Based on the recognized social cues, the Scenario Manager chooses an appropriate reactive behavior model. The Behavior Manager transforms this model into a sequence of timely aligned multimodal character control commands (e.g. speech, gestures, facial expressions, eye and head movement) which are then executed by the character rendering environment.

4.1 Scenario Manager - Behavioral Modeling

For modeling the behavior of our 2 interactive virtual recruiters, we rely on the VisualSceneMaker authoring tool [13] that allows to model and to execute behavioral aspects at both a very detailed and an abstract level.

Central to this tool is the separation of dialog content and interaction structure, see Fig. 2. With VisualSceneMaker, authors work with a visual IDE to specify dialogue behavior in a *scenescript* and to model the dialogue flow with a *sceneflow*, a state chart dialect which is specialized for this purpose. These sceneflows allow the hierarchical refinement and the parallel decomposition of the behaviour model. Thus, we were able to model parallel processed for user input processing and behaviour generation that were properly synchronized.

The multimodal dialog content is specified with a number of scenes that are organized in a scenescript, see Fig. 2, lower half. The scene structure can be compared to those in TV or theater playbooks consisting of utterances and stage directions for the actors. In our scenes, directions are animation commands for gestures, facial expressions, or posture changes. The interactive recruitment simulation is controlled by several parallel hierarchical finite state machines (HFSM) specifying: 1) the logic which scenes are played in which utterance and gesture commands are executed, 2) the characters' gaze and head tilt behavior, and 3) the turn-taking behavior, see Fig. 2, upper half. The turn-taking behavior employs the voice activity social cue of the current user. If the user is speaking, the job interviewer is listening. However, if the job interviewer speaks (e.g. asks the user a question) the user does not have the possibility to barge in.

The modeling of the head tilt and gaze behavior follows the finite state machine implementation of the iGaze [18] approach. But instead of a submissive gaze strategy, the the friendly gaze approach by by Fukayama et al. [12] was used. The parallel executed head tilt and gaze automata uses the social cues (e.g. audio features like voice activity) to change the current state of the virtual character's gaze, see Fig. 3. This automaton consists of 4 major states, defining speaking and listening behavior for 1) the dominant (cf. demanding) character, and 2) the friendly (cf. understanding) character. Each state models a different head tilt and gaze behavior. For example, see Fig. 3 (lower part), the speak friendly HSFM consists of 2 parallel sub automata: 1) one defining the related head tilt movements, 2) another defining the related gaze behavior. The friendly gaze behavior is defined by a longer look at user time, between 3 and 5 seconds, compared to the avert user gaze time, about 0.5 to 1 second. The numbers of the gaze functions define the minimal and maximal degree of eye movement. The friendly head tilt behavior is defined by 3 states, 1) one letting the character tilt its head to the left, 2) tilt the head to the right, and 3) no head tilt. The head tilt time is between 0.5 and 1.5 seconds. While listening, we modeled that the understanding character does not avert the gaze of the user whereas the demanding character might avert her gaze.

The demanding characters show a different head tilt and gaze behavior. While speaking and listening, it shows fewer head tilt and the dominant gaze behavior described in the iGaze approach (e.g. outstaring and sometimes condescending avert of gaze). See Fig. 1 for some symptomatic head tilt and gaze situations during the interaction.

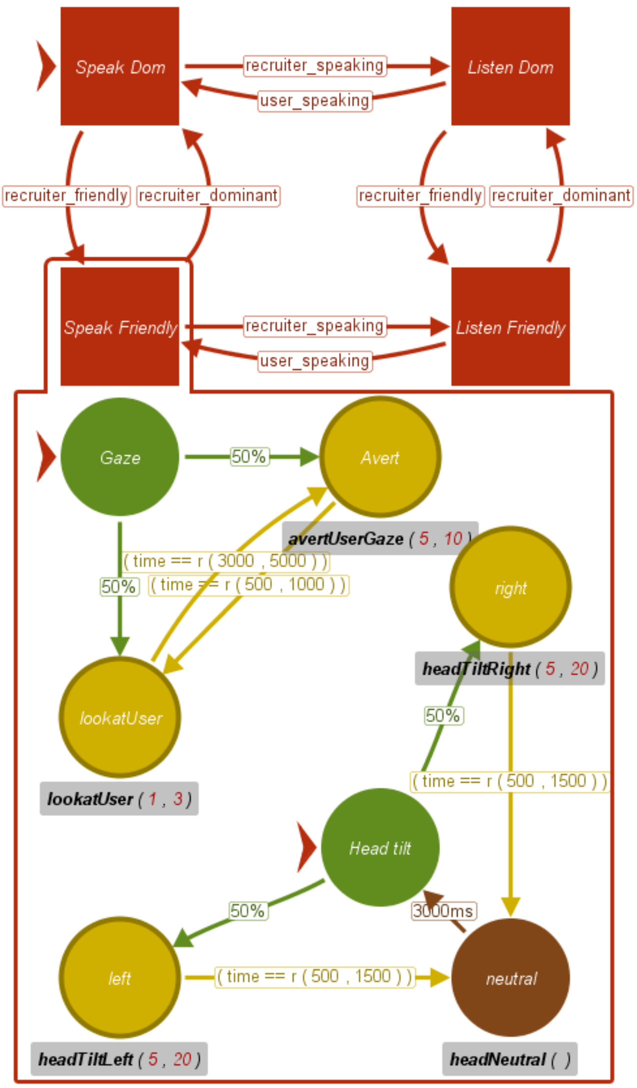

Figure 3: Head tilt and gaze model as HSFM.

The story and interaction model is the major part of virtual recruiter behavior model. Compared to a linear theater scene play, an interactive presentation comes along with another degree of freedom: the reactions of the system on user input.

Fig. 4 shows some exemplary parts of the sceneflow that we used for the evaluation in this paper (left side). The right side shows a few scenes examples. All scenes are created by an author relying on the verbal and non-verbal rules, defined in Sec. 3. A comparison between the scenes for the demanding character with the scenes for the understanding character, reveals that these scenes show a polite language (e.g "*please*", "*would you ...*"), shorter pauses, en-

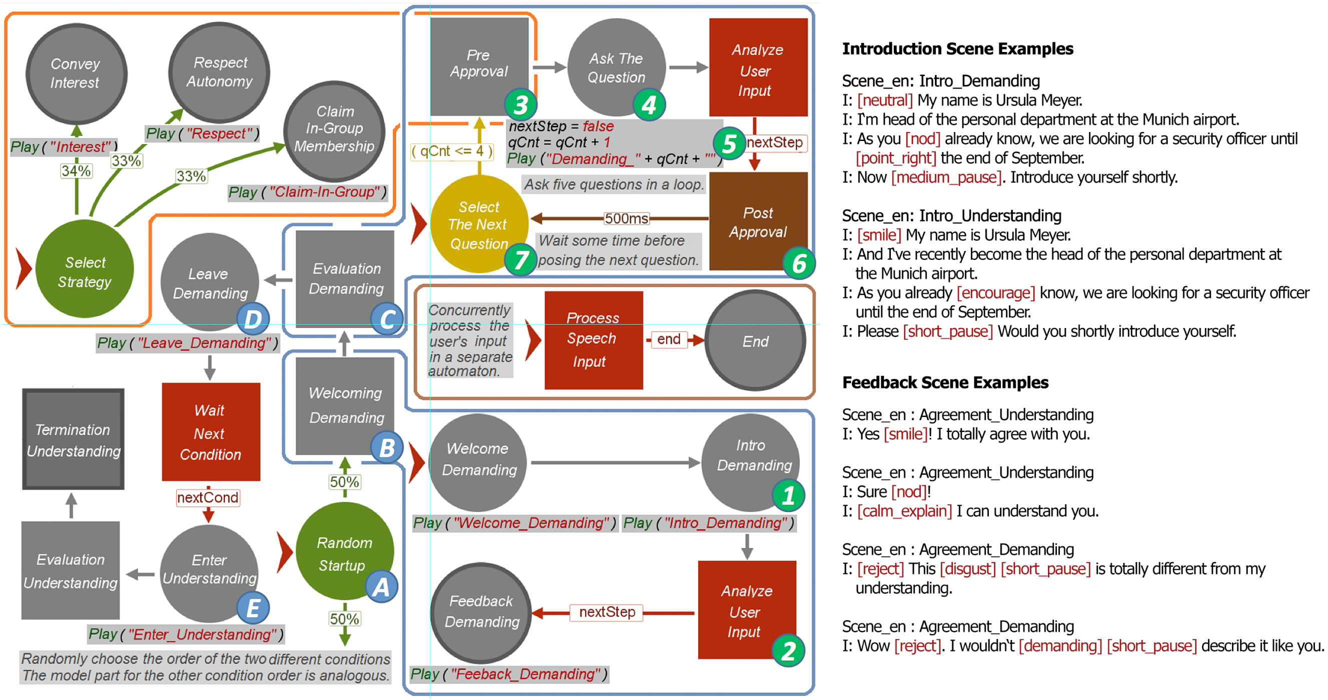

Figure 4: A collage of the HFSM interaction structure defining the characters' dialog and utterance behavior.

couraging gestures (e.g. [*encourage*] or [*calm_explain*]) with a smaller spatial extension, and positive facial expressions (e.g. [*smile*]). Scenes for the demanding character show a formal language, longer pauses, gesture with a larger spatial extension (e.g. [*point_right*]), and negative facial expressions (e.g. [*disgust*] *or* [*demanding*]).

The dialogue starts with a randomly selected order of the two conditions, the demanding and the understanding job interviewer (Fig. 4 Ⓐ). The dialogues in both conditions have the same structure. First each dialogue starts with a welcome phase (Fig. 4 Ⓑ). In this phase the character first welcomes the user and introduces herself to the user (Fig. 4 ①). Afterwards the character asks the user to introduce himself to her (Fig. 4 ②). In a second phase the character poses a set of questions (Fig. 4 Ⓒ). Each question is preceeded by a pre approval phase (Fig. 4 ③) in which the virtual job recruiter introduces the question with some statements following a randomly chosen strategy, e.g *ConveyInterest*, *RespectAutonomy*, or *ClaimInGroupMembership* (see Sec. 3). This phase is immediately followed by the actual question to the user (Fig. 4 ④). Afterwards the character waits for the user's answer (Fig. 4 ⑤). During the user speaks, the recruiter listens and the automata for the listening behavior (head tilt and gaze) (see Fig. 3) becomes active. As soon as the Social Cue Recognizer detects a significant period of no user voice activity, the post approval phase (Fig. 4 ⑥) is executed in which the virtual job recruiter comments the user's answer with some statements following a randomly chosen strategy, e.g *Agreement*, *Feedback*, or *Autonomy* (see Sec. 3). Then, the next question is selected from the scenescript (Fig. 4 ⑦). After five questions have been evaluated, the current character leaves the scene (Fig. 4 Ⓓ) and the next condition starts (Fig. 4 Ⓔ).

The VisualSceneMaker relies on a multi-threaded Java™ interpreter that executes the model and, thus, generates the character's behavior and controlling during the interaction.

The IDE enfolds a runtime visualization mechanism that highlights scenes, states and transitions that are executed. This mechanism facilates iterative prototyping and model testing. The VisualSceneMaker's plug-in mechanism allowed us to easily integrate Charamel's 3D character engine as well as the Social Cue Recognition sensor pipeline.

4.2 Social Cue Recognition

To facilitate seamless interaction between the user and the virtual character, our system uses the SSI framework [27] to analyze various social signals. SSI provides an interface to a large diversity of sensing devices as well as a variety of tools for the real time recording and pre-processing human behavior data. In this work, we use a Microsoft Kinect and a headset. These two sensors give us access to various movement and audio features while having a low intrusion factor. Intrusion is a critical element in this scenario as highly intrusive sensors can artificially influence the users' state of mind which may in turn affect their performance during the interview simulation. For example, biological signal sensors as in the work by Prendinger and Ishizuka [23] are not feasible in this scenario because attaching various sensors to the skin of the users will most likely result in an increase in stress. Therefore, in the context studied, remote sensors are preferred. Our system is able to detect the following cues:

- *Body and Facial Features:* Postures, gestures, head gaze, smiles, motion energy, overall activation
- *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 character's actions on the user.

The movement social cues are implemented using algorithms developed as part of previous work [3]. In order to compute the audio features intensity, loudness, pitch and energy we use OpenSMILE [11]. More complex features are calculated by PRAAT [7] functionality that has been integrated into our system.

4.3 Behavior Analysis Tool

The SSI framework also enables the recording of the social signals so that they can be analyzed at a later point. The NOnVerbal behavior Analyzer (NOVA) is used to analysis of recorded job interview simulations [4]. Through intelligent visualization of social signals, NOVA enables the detailed review of the interactions during an interview, displaying both the actions and social cues of the virtual character and the user.

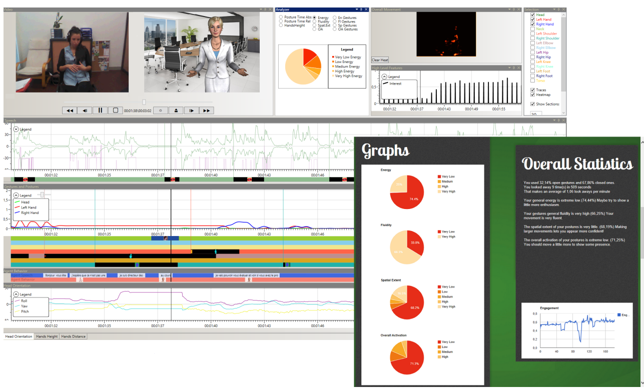

Figure 5: The NOnVerbal behavior Analyzer tool.

The tool is meant to be used by professional practitioners to review the interviews of their pupils as well as by the users themselves in order to enhance the learning process through refection. Additionally, the tool summarizes the interview sessions using various measures computed according to literature. These measures give a user the opportunity to quickly asses her performance and also track the performance over different sessions. For our purposes we used the tool to find differences in the body language and voice of the users in both conditions.

5. PRELIMINARY EVALUATION

First, we are interested whether the users notice the difference between the 2 characters at all. Secondly, we investigate the impact that the characters might have on the users' perception of the learning experience. Thirdly, we aimed to evaluate whether the two characters had an effect on the user's interaction behavior.

5.1 Design and Procedure

Each participant was seated at a table in front of a 24 inch display at a distance of 1.5m. Above the display, a Microsoft Kinect was positioned facing the participant (see Fig. 6). After a quick introduction, the participants provided demographic data. They then conducted a simulated job interview with the two characters. The experiment was designed as a within-subjects experiment. That is all participants interacted with both characters while clothing (white and gray) and the order of their appearance was randomized to avoid any bias due to habituation effects.

The simulated job interview conducted with both characters was structured into 5 parts: Welcome, Introduction, Elaboration, Synthesis and Leave-Taking. Each job interview consisted a total of 10 questions.

Figure 6: The experiment setting.

The first recruiter went through the questions one by one applying turn taking as described in section 4.1. After that, the virtual character left under a pretense. During this break, the participants filled in a questionnaire to indicate how they perceived the character (6 questions) and how they experienced the interview (6 questions). Once they filled in the questionnaire, the second recruiter came in to continue the interview. The procedure was exactly the same as for the first interview. However, the questions were slightly varied in order to avoid habituation effects and maintain realism. Upon completion of the interview, the participants filled in a second questionnaire identical to the first.

5.2 Results and Discussion

A total of 24 participants, 7 female and 17 male with an average age of 28.71 took part in the study. The questionnaire data was evaluated using two sided paired t-tests. We achieved significant results for all values except the virtual characters' question difficulty (see Fig. 7). Overall, the questionnaire data shows that the 2 characters were perceived in the intended manner. Furthermore, they revealed that the personality profiles did have an impact on user experience. Our participants had the feeling that the demanding character induced a higher level of stress in them than the understanding character. They also felt less comfortable when interacting with the demanding character and perceived the interview with this character as more challenging. Even though they did not have the impression that the questions of the demanding character were more difficult, they rated their own performance lower when interacting with this character. It was also harder for them to get immersed into the situation of a job interview when interacting with the demanding character, which was also perceived as less natural than the understanding character. We assume that the demanding character did not match the particpants' expectations of a typical job recruiter who should be friendly and supportive in their mind.

Overall, the approach succeeded in conveying a completely different learning atmosphere by employing two opposed character personalities. The fact that the question difficulty item did not yield significant differences suggests that even though the questions of the two recruiter were not exactly the same, they were similar in difficulty and did not influence the user experience.

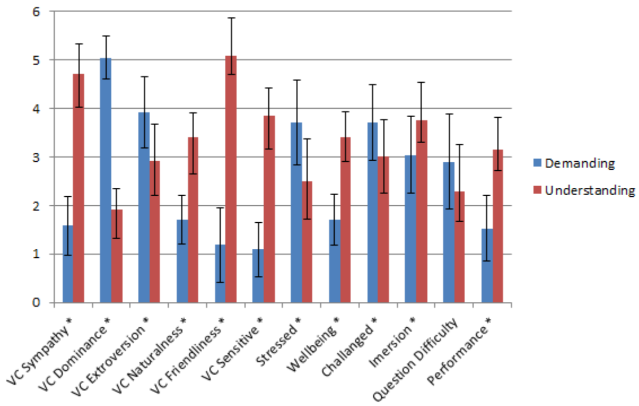

Figure 7: Results of the questionnaire evaluation. Starred items note significant differences (p<0.05).

The audio feature analysis shows that the users performed more breathing pauses in their speech during the understanding condition (2.72 utterances per turn) than in the demanding condition (2.08 utterances per turn), $p = 0.038$. This suggests that the users were more relaxed during the understanding condition, feeling less pressured to give fast answers. These results are backed up by the fact that in the demanding condition the users had on average longer speech chunks ($mean = 3.00$) than in the understanding condition ($mean = 2.62$), $p = 0.015$.

Upon analysis of the movement features we were also able to find some interesting trends albeit not significant. The movement energy measured during the demanding condition was higher (54% of all movements have been categorized as highly energetic while 28% were had a low or very low energy value) than in the understanding condition (39% high or very high energy and 49% low or very low energy movements). This might be related to the reported increased stress level for the demanding condition (see Fig. 8).

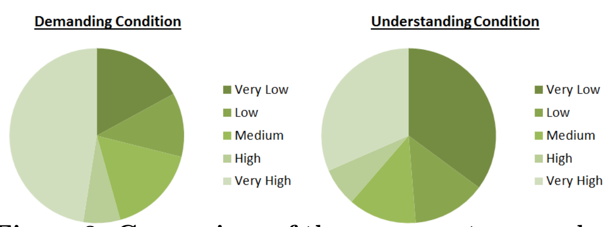

Figure 8: Comparison of the movement energy level during the two conditions.

Overall, the approach succeeded in modulating the learning experience by the two character personality profiles in the intended manner. Furthermore, the character personalities had an impact on the users' performance in the job interview simulation. Thus an important prerequisite is fulfilled to create a learning environment that exposes users to social situations of increasing difficulty.

6. CONCLUSIONS AND FUTURE WORK

In this paper we investigated the influence of contrary interaction strategies for a virtual recruiter on interviewees. Based on previously researched theories about politeness, back channeling and dominance we created two social behavior profiles. One character used an understanding strategy to approach the interviewees during the job interview simulation, while the other character behaved way more demanding. We conducted questionnaires right after each condition and found significant differences within the perception of both characters. In general, participants stated that they felt a higher amount of stress when interacting with the demanding character. Further analysis of the social cues, such as audio features and body language analysis also confirmed this for most participants.

Obviously, the behavior of the character did have an impact on the participants' subjective experience and their performance in the simulation. Our future work will concentrate on the development of additional character personalities. By engaging in role play with characters, job seekers may learn how to adapt to challenging social situations while behavior analyses modules enable them to compare their behaviors under different conditions.

7. ACKNOWLEDGMENTS

This work was partially funded by the European Commission within FP7-ICT-2011-7 (Project TARDIS, grant agreement no. 288578). We thank the people from Charamel GmbH for their continuous support and for providing us with the virtual character Gloria.

8. REFERENCES

[1] E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogues for animated presentation teams. In J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors, *Embodied Conversational Agents*, pages 220–255. MIT Press, 2000.

[2] L. M. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer. Cicero - towards a multimodal virtual audience platform for public speaking training. In R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, editors, *Intelligent Virtual Agents - 13th International Conference, IVA 2013, Edinburgh, UK, August 29-31, 2013. Proceedings*, volume 8108 of *Lecture Notes in Computer Science*, pages 116–128. Springer, 2013.

[3] T. Baur, I. Damian, P. Gebhard, K. Porayska-Pomsta, and E. André. A job interview simulation: Social cue-based interaction with a virtual character. In *Proceedings of the IEEE/ASE International Conference on Social Computing (SocialCom 2013)*, pages 220–227. IEEE, Washington D.C., USA, 2013.

[4] T. Baur, I. Damian, F. Lingenfelser, J. Wagner, and E. André. Nova: Automated analysis of nonverbal signals in social interactions. In *Human Behavior Understanding*, pages 160–171. Springer, 2013.

[5] N. Bee, S. Franke, and E. Andre. Relations between facial display, eye gaze and head tilt: Dominance perception variations of virtual agents. In *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII)*, page 7 pages. IEEE, 2009.

[6] N. Bee, C. Pollock, E. André, and M. A. Walker. Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors. In J. M.

Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, and A. Safonova, editors, *Intelligent Virtual Agents, 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings*, volume 6356 of *Lecture Notes in Computer Science*, pages 265–271. Springer, 2010.

[7] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. version 5.3.56, retrieved 15 september 2013, 2013.

[8] P. Brown and S. C. Levinson. *Politeness — Some universals in language usage*. Cambridge University Press, Cambridge, 1987.

[9] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6:169 – 200, 1992.

[10] B. Endraß, E. André, M. Rehm, and Y. I. Nakano. Investigating culture-related aspects of behavior for virtual characters. *Autonomous Agents and Multi-Agent Systems*, 27(2):277–304, 2013.

[11] F. Eyben, M. Wöllmer, and B. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the international conference on Multimedia*, MM '10, pages 1459–1462, New York, NY, USA, 2010. ACM.

[12] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents — impression management with agent's gaze. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 41–48, New York, NY, USA, 2002. ACM.

[13] P. Gebhard, G. Mehlmann, and M. Kipp. Visual scenemaker—a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces*, 6(1-2):3–11, 2012.

[14] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, editors, *Intelligent Virtual Agents, 7th International Conference, IVA 2007, Paris, France, September 17-19, 2007, Proceedings*, volume 4722 of *Lecture Notes in Computer Science*, pages 125–138. Springer, 2007.

[15] J. O. Greene and B. R. Burleson. *Handbook of Communication and Social Interaction Skills.* Lawrence Erlbaum Associates, Mahwah, New Jersey, London, 2003.

[16] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard. MACH: my automated conversation coach. In F. Mattern, S. Santini, J. F. Canny, M. Langheinrich, and J. Rekimoto, editors, *The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13, Zurich, Switzerland, September 8-12, 2013*, pages 697–706. ACM, 2013.

[17] S.-H. Kang, J. Gratch, C. L. Sidner, R. Artstein, L. Huang, and L.-P. Morency. Towards building a virtual counselor: modeling nonverbal behavior during intimate self-disclosure. In W. van der Hoek, L. Padgham, V. Conitzer, and M. Winikoff, editors, *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4-8, 2012 (3 Volumes)*, pages 63–70. IFAAMAS, 2012.

[18] M. Kipp and P. Gebhard. IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In H. Prendinger, J. C. Lester, and M. Ishizuka, editors, *Intelligent Virtual Agents, 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings*, volume 5208 of *Lecture Notes in Computer Science*, pages 191–199. Springer, 2008.

[19] R. MacDonald. Disconnected youth? social exclusion, the underclass and economic marginality. *Social Work and Society*, 6(2):236–247, 2008.

[20] F. Mairesse and M. A. Walker. Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Model. User-Adapt. Interact.*, 20(3):227–278, 2010.

[21] M. Neff, Y. Wang, R. Abbott, and M. A. Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *Intelligent Virtual Agents, 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings*, volume 6356 of *Lecture Notes in Computer Science*, pages 222–235. Springer, 2010.

[22] R. Niewiadomski, J. Hofmann, J. Urbain, T. Platt, J. Wagner, B. Piot, H. Cakmak, S. Pammi, T. Baur, S. Dupont, M. Geist, F. Lingenfelser, G. McKeown, O. Pietquin, and W. Ruch. Laugh-aware virtual agent and its impact on user amusement. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, AAMAS '13, pages 619–626, Richland, SC, 2013. International Foundation for Autonomous Agents and Multiagent Systems.

[23] H. Prendinger and M. Ishizuka. The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence*, 19(3-4):267–285, 2005.

[24] B. Ravenet, M. Ochs, and C. Pelachaud. *From a User-created Corpus of Virtual Agent's Non-verbal Behavior to a Computational Model of Interpersonal Attitudes*, volume 8108, pages 263–274. Springer Berlin Heidelberg, 2013.

[25] M. Schröder. The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems. *Advances in Human-Computer Interaction*, 2010, 1 2010. http://www.odysci.com/article/1010113017465674.

[26] D. R. Traum, D. DeVault, J. Lee, Z. Wang, and S. Marsella. Incremental dialogue understanding and feedback for multiparty, multimodal conversation. In Y. Nakano, M. Neff, A. Paiva, and M. A. Walker, editors, *Intelligent Virtual Agents - 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*, volume 7502 of *Lecture Notes in Computer Science*, pages 275–288. Springer, 2012.

[27] 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 *Proceedings of the 21st ACM International Conference on Multimedia, 21-25 October 2013, Barcelona, Spain.*, 2013.