PARLEY:
A Transparent Virtual Social Agent Training Interface

Tanja Schneeberger
Patrick Gebhard
firstname.lastname@dfki.de
German Research Center for Artificial Intelligence
Saarbrücken, Germany

Tobias Baur
Elisabeth André
lastname@hcm-lab.de
Augsburg University
Augsburg, Germany

ABSTRACT

In this demo, we present PARLEY, an interactive system to train difficult social situations in a safe environment with a Virtual Agent. The system realizes different phenomena studied by psychology research that are known to create a natural interaction. Moreover, we include an open learner model to ensure an explainable user experience.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Interactive systems and tools**.

KEYWORDS

Social Training System, Explainable AI, Virtual Agents

1 INTRODUCTION

As social beings, humans are living in social communities at home and at work, being confronted with familiar social situations but also with uncertain new scenarios. Especially, rare, unfamiliar or emotionally difficult situations are challenging for most of the people. There is no way to prepare for these social situations interactively in a protected room without pressure. Therefore, we created PARLEY, a social training system that combines interactive HCI technology with psychological models to build novel agents that are capable of explaining their decisions. This way, we create an explainable user interface to increase users' trust and comprehensibility of the system's decisions. In this demo, we present PARLEY as it can be exploited for the use case of a job interview training. As the most common selection procedure [7], a job interview is a highly evaluative and therefore difficult situation [5].

Figure 1: The PARLEY system including the Virtual Agent, the depth camera and the screen showing the open learner model.

2 SYSTEM OVERVIEW

The system is based on a combination of the Social Signal Interpretation (SSI) Framework [10] and the hybrid authoring tool VisualSceneMaker (VSM) [3]. SSI offers the possibility to extract multimodal behavioral characteristics and to combine them by fusion methods to a holistic analysis. SSI exploits the latest AI technologies e.g., Deep Learning for the classification of paralinguistics and other social signals. The results of this holistic analysis are made available to VSM which is equipped with a real-time execution component (latency time 25-50ms), and it controls the agents' behavior (e.g., gestures, speech). VSM integrates autonomous and reactive behavior with learnable behavior sequences. With VSM, the behavior and interaction management, the dialog flow, and the content are modeled. Consequently, the content of the interaction with the system is similar for every user. This ensures that the system can be used for social trainings as well as for user studies. It makes it possible to vary details in the interaction or the content and to examine those differences in controlled experiments. Concerning real-time reactive nonverbal behavior, VSM lets authors specify behavior rules that enable the PARLEY agents to react naturally. Additionally, it is possible to control the agents behavior in Wizard of Oz studies via a remote control.

3 INTERACTION DESIGN

Various aspects (e.g., Interruptions, Mimicry) that make human-human-interactions appear natural are implemented in the PARLEY human-agent-interaction The primary goal is to create interactive social agents that demonstrate credible social communication behavior. Most importantly, the system does not use any explicit

input devices such as a keyboard or a mouse. Neither equipping the user with disturbing sensors is necessary. The user can freely approach the system, it recognizes him and initiates the interaction. In order to ensure a natural flow of conversation, the system recognizes when the user stops speaking and gives the turn to the Virtual Agent. Through a broad repertoire of facial expressions and gestures, the emotional expression of the Virtual Agent can be realistically designed, giving the impression that the agent might be able to build a relationship with the user.

3.1 Interruptions

Though there is work examining interruptions in the area of Virtual Agents [1], none of the systems presented the possibility to interrupt a Virtual Agent. However, this enables the user to ask questions in case of ambiguities or to correct erroneous assumptions. The PARLEY agents are designed to respond to interruptions by the user by stopping and resuming their utterance at the right moment.

3.2 Mimicry

Mimicry describes the phenomenon that people unconsciously and automatically imitate other people in interactions and social situations. As a relationship regulator, mimicry leads to social proximity when applied and to social coldness when not applied [6]. The PARLEY agents can mirror the behavior of the user and thus use mimicry for positive relationship regulation, whereby the user gets the feeling of interacting with an empathetic counterpart. For training purposes, mimicry can be also disabled to have opposite effects.

3.3 Backchanneling

Backchanneling occurs when one of the speakers of a dialogue pair speaks, and the other focuses on listening while giving non-intrusive acoustic and visual signals [11]. It provides i.a. information about the listeners attention and comprehension [8]. The PARLEY agents have been designed to show backchanneling while the user is speaking. Smile and nod, as well as short utterances like "Hm", "Yes", "Okay", signal to the user that he is being listened to attentively.

3.4 Emotion Model

The basis for successful dyadic communication, in addition to the behavioral aspects mentioned, is an understanding of the inner experience of the other. In order to conclude about what is going on in a user, it is not enough to analyze only the visible social signals. A smile, for example, can not only be an expression of joy but can also be used to cover up insecurity. PARLEY employs the MARSSI emotion model [4], a multi-layered emotion model combining a real-time analysis of social signals with an analysis of the context and the consideration of individual emotion regulation strategies. MARSSI uniquely employs the classification between structural, communicative and situational emotions, what allows the differentiation between external (communicative) and internal (structural and situational) emotions and their relation based on emotion regulation strategies. Therefore, MARSSI aims to, e.g., differentiate between a smile out of happiness or a smile that should cover insecurity in an inconvenient situation. The PARLEY system simulates internal emotions based on communicated and captured signals in a Bayesian probability space. This emotion model enables

the PARLEY agents to show behavior that is not only related to the superficially recognizable user behavior but also the internal experience of the user.

3.5 Explainable AI: Open Learner Model

At the end of the interaction, users are provided with an overview of their reactions and their interpretation by the system in order to reflect objectively and transparently on their appearance (Open Learner Model). The transparent visualization strengthens the trust in the system and thus increases the learning effect [2]. For this purpose, we apply techniques from the current research area of Explainable AI [9].

4 CONCLUSION

In this demo, we introduce PARLEY an interactive system that enables users to practice social situations. By combining research from the field of emotion psychology and social signal interpretation, we create a novel interaction possibility that is based on an explainable user interface. Thus, we give users the possibility to practice emotionally difficult social situations in a protected space.

ACKNOWLEDGMENT

REFERENCES

[1] Angelo Cafaro, Nadine Glas, and Catherine Pelachaud. 2016. The effects of interrupting behavior on interpersonal attitude and engagement in dyadic interactions. In *Proceedings of the 2016 Int. Conference on Autonomous Agents & Multiagent Systems.* 911–920.

[2] Ionut Damian, Tobias Baur, Birgit Lugrin, Patrick Gebhard, Gregor Mehlmann, and Elisabeth André. 2015. Games are better than books: In-situ comparison of an interactive job interview game with conventional training. In *Int. Conference on Artificial Intelligence in Education.* Springer, 84–94.

[3] Patrick Gebhard, Gregor Mehlmann, and Michael Kipp. 2012. Visual scenemaker - a tool for authoring interactive virtual characters. *Journal on Multimodal User Interfaces* 6, 1-2 (2012), 3–11.

[4] Patrick Gebhard, Tanja Schneeberger, Tobias Baur, and Elisabeth André. 2018. MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation. In *Proceedings of the 17th Int. Conference on Autonomous Agents and Multiagent Systems.* Int. Foundation for Autonomous Agents and Multiagent Systems, 497–506.

[5] Richard G. Heimberg, Kevin E. Keller, and Theresa A. Peca-Baker. 1986. Cognitive assessment of social-evaluative anxiety in the job interview: Job Interview Self-Statement Schedule. *Journal of Counseling Psychology* 33, 2 (1986), 190.

[6] Ursula Hess and Agneta Fischer. 2013. Emotional mimicry as social regulation. *Personality and Social Psychology Review* 17, 2 (2013), 142–157.

[7] Julia Levashina, Christopher J. Hartwell, Frederick P. Morgeson, and Michael A. Campion. 2014. The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology* 67, 1 (2014), 241–293.

[8] Christopher Peters, Catherine Pelachaud, Elisabetta Bevacqua, Maurizio Mancini, and Isabella Poggi. 2005. A model of attention and interest using gaze behavior. In *Int. Workshop on Intelligent Virtual Agents.* Springer, 229–240.

[9] Johannes Wagner, Tobias Baur, Dominik Schiller, Yue Zhang, Björn Schuller, Michel Valstar, and Elisabeth André. 2018. Show Me What You've Learned: Applying Cooperative Machine Learning for the Semi-Automated Annotation of Social Signals. *IJCAI/ECAI, Workshop on XAI 2018* (2018), 171–177.

[10] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM Int. Conference on Multimedia.* ACM, 831–834.

[11] Victor H. Yngve. 1970. On getting a word in edgewise. In *Chicago Linguistics Society, 6th Meeting, 1970.* 567–578.