SimSensei Demonstration: A Perceptive Virtual Human Interviewer for Healthcare Applications

Louis-Philippe Morency, Giota Stratou, David DeVault, Arno Hartholt,
Margaux Lhommet, Gale Lucas, Fabrizio Morbini, Kallirroi Georgila,
Stefan Scherer, Jonathan Gratch, Stacy Marsella, David Traum, Albert Rizzo
Institute for Creative Technologies, University of Southern California
Los Angeles, California

Abstract

We present the SimSensei system, a fully automatic virtual agent that conducts interviews to assess indicators of psychological distress. We emphasize on the perception part of the system, a multimodal framework which captures and analyzes user state for both behavioral understanding and interactional purposes.

Introduction

Virtual humans that can develop intimacy with people are now becoming reality. Researchers have successfully incorporated social skills (e.g., active listening, mimicry, gestures) into virtual human systems (Bickmore, Gruber, and Picard 2005; Gratch, Kang, and Wang 2013). Indeed, compared to their predecessors, virtual humans with such social skills increase feelings of connection and rapport, namely the experience of harmony, fluidity, synchrony, and flow felt during a conversation; (Gratch, Kang, and Wang 2013). Enhanced with these skills, virtual humans could be particularly useful as tools. For example, a user experience can be better standardized with virtual humans than with human beings. Recently it was also shown that virtual humans can increase willingness to disclose by providing a "safe" environment where participants don't feel judged by another human interlocutor (Lucas et al. 2014). This is particularly helpful in the case of healthcare scenarios where disclosure of sensitive information is prominent. An automatic system behind the virtual human brings other advantages as well. Utilizing advances in natural language understanding, facial expression analysis and gesture recognition scientists have the tools to analyze and quantify verbal and nonverbal behavior exchanges during real-time interactions. This capability brings forward new opportunities in applications where measuring and assessing human behavior is important (for example in healthcare).

With this demonstration, we present the SimSensei system, a fully automatic system designed to conduct interviews related to psychological distress (DeVault and et al. 2014). In the live demo we emphasize on the perception part of the system, a multimodal framework which captures and analyzes user state for both behavioral understanding and interactional purposes.

Figure 1: This is a side by side view of a participant interacting with Ellie, the SimSensei agent (seen on the right side). On the left side, we see the participant being tracked in real-time by MultiSense, the multimodal perception system used in SimSensei. The participant's tracked state informs the automatic system and enhances the interaction.

SimSensei System Overview

The SimSensei system, seen in Fig.1 is a fully automatic framework designed to create an engaging face-to-face interaction with a user. It was developed as a clinical decision support tool that would complement existing self-assessment questionnaires by giving healthcare providers objective measurements of the user behaviors that are correlated with psychological distress. As such, the main design goals were to create an interactional environment that allows the user to feel comfortable talking during a semi-structured interview, so that elicited behaviors can be automatically measured and analyzed under the specific context.

The SimSensei system has been designed and developed over a two year iterative process described in detail in (DeVault and et al. 2014). It is implemented using a modular approach (Hartholt et al. 2013) and its main components include a dialogue processing component, supported by individual modules for speech recognition, language understanding and dialogue management, a nonverbal behavior generation component for the virtual human, a rendering engine and a multimodal perception system that analyzes au-

diovisual streams in real time.

Perception System Overview

The perception system was tailored specifically for this application (details on the iterative process can be found in (DeVault and et al. 2014)). Specifically, it should serve a double purpose: i) communicate the necessary nonverbal behavior signals to the other components of the system so that the agent is sensitive to the user's nonverbal behavior, and ii) recognize automatically and quantify the nonverbal behaviors. As an example, tracking the smile intensity of the participant serves both of these purposes: smile is a signal that has been tied to investigations of depression and also plays an important role in a dyadic interaction. As a basis for the perception system, SimSensei uses the MultiSense framework, which is a flexible system for multimodal real-time sensing. MultiSense allows for synchronized capture of different modalities, and provides a flexible platform for real-time tracking and multimodal fusion. MultiSense dynamically leverages the measures from current technologies into informative signals such as smile intensity, 3D head position, intensity or lack of facial expressions like anger, disgust and joy, speaking fraction, gaze direction etc. As mentioned above, these informative signals serve two purposes. First, they contribute to the indicator analysis. Second, they are broadcasted to the other components of SimSensei system using the PML standard (Scherer et al. 2012) to inform the virtual human of the state of the participant and assist with turn taking, listening feedback, and building rapport by providing appropriate non-verbal feedback.

Evaluation

With respect to providing an environment where informative behaviors can be elicited and analyzed in relation with distress, relevant work (DeVault et al. 2013; Scherer et al. 2013; Stratou et al. 2014) on the verbal channel and on audio and video modalities respectively, showed that a virtual human interaction is still rich in behaviors that allow for analysis and assessment of distress via fully automatic methods. This was still in the wizard-driven phase of the system, but similar results can be reproduced in the fully automatic system (different stages of the system and the data we collected are described in (Gratch et al. 2014)).

Moreover, it was possible to assess the rapport the users felt with our system through self-report questionnaires in each phase. Overall the results were promising; participants reported willingness to disclose, willingness to recommend and general satisfaction with both the wizard-driven and the fully AI version of the system (Lucas et al. 2014). Also, there is no significant difference in felt rapport as reported by the users between face-to-face interactions with another human and the final AI system, as discussed in (DeVault and et al. 2014). Interestingly, participants felt more rapport when interacting with the wizard-driven system than they did in face-to-face interviews. This is further explored with the fully AI system showcasing that even with the same virtual human interface, the belief that participants interact with a fully automatic system versus a human operator behind the scenes increases their willingness to disclose (Lucas et al. 2014). These results suggest that automated virtual humans that can engage participants to disclose information can help overcome a significant barrier to obtaining truthful patient information

Acknowledgments

This work is supported by DARPA under contract (W911NF-04-D-0005) and U.S. Army Research, Development, and Engineering Command. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

Bickmore, T.; Gruber, A.; and Picard, R. 2005. Establishing the computerpatient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59(1):21 – 30.

DeVault, D., and et al. 2014. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '14, 1061–1068. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

DeVault, D.; Georgila, K.; Artstein, R.; Morbini, F.; Traum, D.; Scherer, S.; Rizzo, A. S.; and Morency, L.-P. 2013. Verbal indicators of psychological distress in interactive dialogue with a virtual human. In *Proceedings of the SIGDIAL 2013 Conference*, 193–202. Metz, France: Association for Computational Linguistics.

Gratch, J.; Artstein, R.; Lucas, G.; Stratou, G.; Scherer, S.; Nazarian, A.; Wood, R.; Boberg, J.; DeVault, D.; Marsella, S.; et al. 2014. *The Distress Analysis Interview Corpus of human and computer interviews*. LREC.

Gratch, J.; Kang, S.-H.; and Wang, N. 2013. Using social agents to explore theories of rapport and emotional resonance. *Social Emotions in Nature and Artifact* 181.

Hartholt, A.; Traum, D. R.; Marsella, S. C.; Shapiro, A.; Stratou, G.; Leuski, A.; Morency, L.-P.; and Gratch, J. 2013. All together now - introducing the virtual human toolkit. In *IVA*, 368–381.

Lucas, G. M.; Gratch, J.; King, A.; and Morency, L.-P. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37(0):94 – 100.

Scherer, S.; Marsella, S. C.; Stratou, G.; Xu, Y.; Morbini, F.; Egan, A.; Rizzo, A.; and Morency, L.-P. 2012. Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In *The 12th International Conference on Intelligent Virtual Agents (IVA)*.

Scherer, S.; Stratou, G.; Gratch, J.; and Morency, L.-P. 2013. Investigating voice quality as a speaker-independent indicator of depression and ptsd. In *Interspeech*, 847–851.

Stratou, G.; Scherer, S.; Gratch, J.; and Morency, L.-P. 2014. Automatic nonverbal behavior indicators of depression and ptsd: the effect of gender. *Journal on Multimodal User Interfaces* 1–13.