

Come and Have an Emotional Workout with Sensitive Artificial Listeners!

Marc Schröder¹, Sathish Pammi¹, Hatice Gunes², Maja Pantic², Michel F. Valstar²,
Roddy Cowie³, Gary McKeown³, Dirk Heylen⁴, Mark ter Maat⁴, Florian Eyben⁵, Björn Schuller⁵, Martin Wöllmer⁵,
Elisabetta Bevacqua⁶, Catherine Pelachaud⁶, and Etienne de Sevin⁶

Abstract—This demonstration aims to showcase the recently completed SEMAINE system. The SEMAINE system is a publicly available, fully autonomous Sensitive Artificial Listeners (SAL) system that consists of virtual dialog partners based on audiovisual analysis and synthesis (see <http://semaine.opendfki.de/wiki>). The system runs in real-time, and combines incremental analysis of user behavior, dialog management, and synthesis of speaker and listener behavior of a SAL character, displayed as a virtual agent. The SAL characters intend to engage the user in a conversation by paying attention to the user’s emotions and nonverbal expressions. The characters have their own emotionally defined personality. During an interaction, the characters attempt to create an emotional workout for the user by drawing her/him towards their dominant emotion, through a combination of verbal and nonverbal expressions.

I. INTRODUCTION

Most of the past research focused on creating virtual agent systems based on static input parameters, rather than dynamically changing the behavior of the virtual agent in accordance with the behavior of the user during an interaction [1], [2]. The SEMAINE system is a pioneering effort in creating dynamic, expressive and adaptive virtual agents by analyzing the multimodal nonverbal communicative behavior of the human user in (soft) real-time.

The system aims to engage the user in a dialog (and create an emotional workout) by paying attention to the user’s nonverbal expressions, and reacting accordingly. It focuses on the ‘soft skills’ that humans naturally use to keep a conversation alive. To simplify the challenge somewhat, the SEMAINE system avoids task-oriented dialog. Instead, it models the type of interaction found at parties: you listen to someone you want to chat with, and without really understanding much of what they are saying, you exhibit all the signs that are needed for them to continue talking to you. The SAL characters can speak to engage the user in a simple dialog as well as show nonverbal listener signals. The approach has been test-run using Wizard of Oz setups at various stages of maturity [3], [4], [6]. This has allowed us to fine-tune the scripts used by the various characters, in order to react to the emotional state of the user in plausible ways despite the lack of language understanding.

II. THE DEMONSTRATION SETUP

During the SEMAINE demonstration, one human user is sitting in front of a computer screen showing the face of an Embodied Conversational Agent (ECA). The user is wearing a headset for voice analysis and is recorded by a video camera for head gesture and facial expression analysis. The ECA is speaking through loudspeakers, and is showing both verbal and nonverbal behavior. A

This work has been funded by the European Community’s 7th Framework Programme [FP7/2007-2013] under grant agreement no 211486 (SEMAINE).

¹ DFKI GmbH, Saarbrücken, Germany; firstname.lastname@dfki.de

² Imperial College London, UK; firstname.lastname@imperial.ac.uk

³ Queens Univ. Belfast, UK; (r.cowie_g.mckeown@qub.ac.uk)

⁴ Univ. Twente, The Netherlands; (d.k.j.heylen_jmaatm@ewi.utwente.nl)

⁵ Technische Univ. München, Germany; lastname@tum.de

⁶ CNRS, Telecom Paristech, France; firstname.lastname@telecomparistech.fr

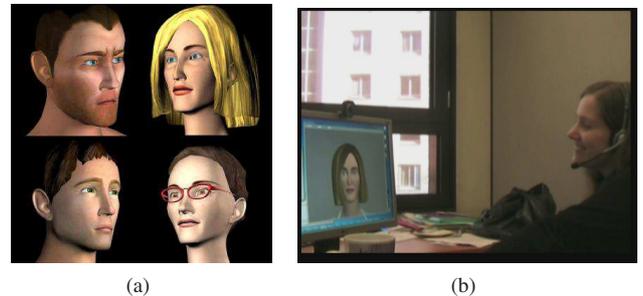


Fig. 1. (a) The four SAL characters represent the four quadrants of arousal-valence space: Spike is aggressive; Poppy is cheerful; Obadiah is gloomy; and Prudence is pragmatic. (b) Illustration of the SEMAINE system: one user conversing with Poppy.

second screen shows a system monitor, displaying graphically the current information flow in the system. The user can speak to one of the four SAL characters at a time (see Fig. 1(a) and Fig. 1(b)).

Each character will try to sustain the conversation by being an active speaker and listener using multimodal verbal utterances and feedback signals. The user can request to switch to a different character whenever (s)he wishes. In the lab, sessions typically last for around 20 minutes; during the demo, much shorter sessions with changing users are anticipated. Technically, the demonstrator system is a multimodal interactive system with components integrated across programming languages and operating systems by means of a standards-based framework for building emotion-oriented systems, the SEMAINE API [5]. Details on the technological setup and the individual processing components are described in a set of project deliverable reports available from the project website: <http://www.semaine-project.eu/>.

We demonstrated the first version of the SEMAINE system at ACII 2009 [6]. This demonstration will showcase the latest and the most refined version of the fully autonomous SEMAINE system released in December 2010.

REFERENCES

- [1] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, “Emotion representation, analysis and synthesis in continuous space: A survey,” in *Proc. of IEEE Int. Conf. on Face and Gesture Recognition*, 2011.
- [2] D. Heylen, M. Theune, R. op den Akker, and A. Nijholt, “Social agents: The first generations,” in *Proc. of ACII*, 2009, pp. 1–7.
- [3] E. Douglas-Cowie, R. Cowie, C. Cox, N. Amir, and D. Heylen, “The sensitive artificial listener: an induction technique for generating emotionally coloured conversation,” in *Proc. of LREC Workshop on Corpora for Research on Emotion and Affect*, 2008, pp. 1–4.
- [4] G. McKeown, M.F. Valstar, R. Cowie, and M. Pantic, “The semaine corpus of emotionally coloured character interactions,” in *Proc. of IEEE ICME*, 2010, pp. 1079–1084.
- [5] M. Schröder, “The semaine api: Towards a standards-based framework for building emotion-oriented systems,” *Advances in Human-Machine Interaction*, vol. 2010, pp. 1–21, 2010.
- [6] M. Schröder et al., “A demonstration of audiovisual sensitive artificial listeners,” in *Proc. of ACII*, 2009, vol. 1, pp. 263–264.