Interacting with emotional virtual agents

Elisabetta Bevacqua, Florian Eyben, Dirk Heylen, Mark ter Maat, Sathish Pammi, Catherine Pelachaud, Marc Schröder, Björn Schuller, Etienne de Sevin, Martin Wöllmer

Interacting with Emotional Virtual Agents

Elisabetta Bevacqua[1], Florian Eyben[3], Dirk Heylen[4], Mark ter Maat[4],
Sathish Pammi[2], Catherine Pelachaud[1], Marc Schröder[2], Björn Schuller[3],
Etienne de Sevin[5], and Martin Wöllmer[3]

[1] CNRS ParisTech, Paris, France
[2] DFKI GmbH, Saarbrücken, Germany
[3] Technische Universität München, Germany
[4] Universiteit Twente, The Netherlands
[5] Université Pierre et Marie Curie, Paris, France
{bevacqua,pelachaud}@telecom-paristech.fr, {eyben,schuller}@tum.de,
{d.k.j.Heylen,maatm}@ewi.utwente.nl,
{Sathish_Chandra.Pammi,marc.schroeder}@dfki.de,
etienne.de-sevin@lip6.fr, woe@mmk.e-technik.tu-muenchen.de

Abstract. Sensitive Artificial Listener (SAL) is a multimodal dialogue system which allows users to interact with virtual agents. Four characters with different emotional traits engage users is emotionally coloured interactions. They not only encourage the users into talking but also try to drag them towards specific emotional states. Despite the agents very limited verbal understanding, they are able to react appropriately to the user's non-verbal behaviour. The demonstrator shows an final version of the fully autonomous SAL system.

1 Introduction

The Sensitive Artificial Listener demo shows the final system developed within the European FP7 SEMAINE project. This project aimed at building a Sensitive Artificial Listener (SAL), a multimodal dialogue system which allows users to interact with virtual agents. The system can sustain an emotionally coloured interaction with users for some time reacting appropriately to the their non-verbal behaviour. The system can perceive user's verbal and non-verbal behaviours and use this information to plan how to react. Its response is transmitted using an ECA that is capable of communicating through several channels like voice, facial expressions, gestures, head movements and torso shifts. The virtual agent not only encourages the user into talking but also try to pull her/him towards specific emotional states. To achieve such a goal, SAL provides four characters with different emotional traits. Spike is nasty, angry creature, and his mission in life is to make the user angry too. Poppy is the happy and positive girl and she tries hard to make her interlocutor as happy as she is. Then there's Prudence. She's sensible and pragmatic about everything and she expects the user to be sensible
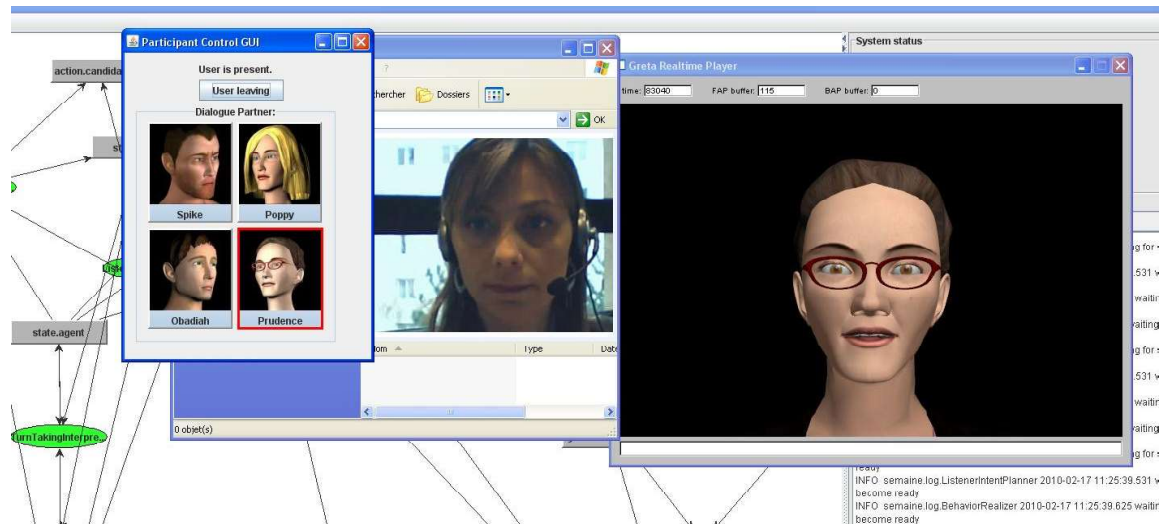
Fig. 1. Demonstration setup

too. Finally, Obadiah is the soul of misery, and he thinks everybody should be miserable, including his interlocutor.

1.1 Demonstration Setup

The demonstration set up is shown in figure 1. The user sits in front of a screen where a SAL character is displayed. The user must wear a microphone for voice analysis and can optionally be recorded by a video camera for facial expression analysis and head movement recognition. The system monitor can be displayed on a second screen; it shows the components and the current data flowing between them. During the interaction the user is the speaker, the virtual agent is mainly the listener who from time to time provides just short sentences to encourages the user into talking. The agent cannot really understand the user's speech, so sometimes the agent's sentences may appear absolutely inconsistent and wild. That is just part of the interaction. It is up to the user to supply all the ideas and the effort to push forward the interaction as much as possible, keeping in mind that there is no point asking the agent questions, or trying to outwit it.

1.2 Technical Description and Requirements

The system uses the SEMAINE API, a distributed multi-platform component integration framework for real-time interactive systems [1]. User's acoustic and visual cues are extracted by analyser modules and then used by the interpreters to derive the system's current best guess regarding the state of the user and the dialogue. This information and the user's acoustic and visual cues are used to generate the agent's behaviour both while speaking and listening.

To run the whole system can be quite cumbersome, so at least a quad core processor machine with 6GB RAM is required. A microphone, web camera and loudspeakers are needed, too. The demo can be easily installed in 10 minutes.

Reference

1. Schröder, M.: The semaine api: Towards a standards-based framework for building emotion-oriented systems. In: Advances in Human-Computer Interaction 2010 (2010)