Greta: an interactive expressive ECA system

Radosław Niewiadomski[1] Elisabetta Bevacqua[1] Maurizio Mancini[2] Catherine Pelachaud[1]

[1]TELECOM ParisTech
46 rue Barrault
75013, Paris - France
radoslaw.niewiadomski, elisabetta.bevacqua,
catherine.pelachaud@telecom-paristech.fr

[2]InfoMus Lab, DIST – University of Genova
via Causa 13
I-16145, Genova - Italy
maurizio@infomus.org

ABSTRACT

We have developed a general purpose use and modular architecture of an Embodied Conversational Agent (ECA) called Greta. Our 3D agent is able to communicate using verbal and nonverbal channels like gaze, head and torso movements, facial expressions and gestures. It follows the SAIBA framework [10] and the MPEG4 [6] standards. Our system is optimized to be used in interactive applications.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: Miscellaneous

General Terms

Algorithms

Keywords

ECA architecture, multimodal behavior, SAIBA, BML, FML

1. INTRODUCTION

In this paper we describe the architecture of an embodied conversational agent (ECA) called Greta. Greta is 3D humanoid agent able to communicate with the user using verbal and nonverbal channels like gaze, facial expressions and gestures. It follows the SAIBA framework [10] that defines a modular structure, functionalities and communication protocols for ECA systems. Moreover, Greta follows the MPEG4 [6] standard of animation. In our system the agent can be at turn a speaker and a listener in conversations between agents or with a human user.

2. GRETA ARCHITECTURE

Figure 1 illustrates the architecture of the Greta agent. Our system uses the FML-APML XML-language [2] to specify the agent's *communicative intentions* (e.g., its beliefs, emotions) that go along with what the agent wants to say. The communicative intentions of the listener are generated by the **Listener Intent Planner** while the intentions of the speaker are defined at the moment manually in an FML-APML input file. In the future they will be generated by the **Speaker Intent Planner**. The **Behavior Planner module** receives as input the agent's communicative intentions written in FML-APML and generates as output a list of signals in BML language. BML specifies the verbal and nonverbal behaviors of

Cite as: Greta: an Interactive Expressive ECA System, Radosław Niewiadomski, Elisabetta Bevacqua, Maurizio Mancini, Catherine Pelachaud, *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009),* Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. 1399 – 1400

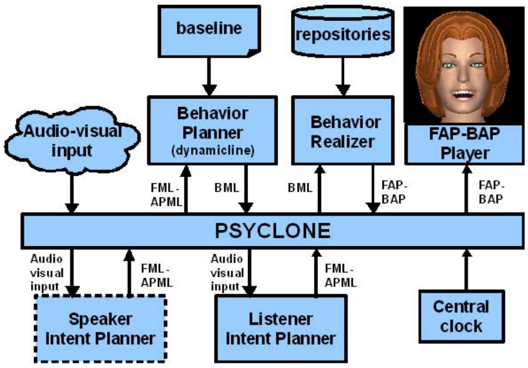

Figure 1. Greta architecture.

the agent [10]. Each BML tag corresponds to a behavior the agent is to produce on a given modality: head, torso, face, gaze, gesture, speech. These signals are sent to the **Behavior Realizer** that generates the MPEG4 FAP-BAP files. Finally, the animation is played in the FAP-BAP Player.

All modules in the Greta architecture are synchronized using a central clock and communicate with each other through a *whiteboard*. For this purpose we use Psyclone messaging system [9] which allows modules and applications to interact through TCP/IP. The system has a very low latency time that makes it suitable for interactive applications.

In the following subsections we describe each module of Greta.

AUDIO-VISUAL INPUT. Video and audio data about the user's verbal and non verbal behavior is needed to know what the user is saying and how s/he is saying it. Such a data is collected and analyzed by external applications using a webcam and a microphone and sent to our system through a TCP/IP connection. Different applications for input analysis can be used and connected with our system without requiring major changes. So far, the system has been connected with Watson [4], a real-time head tracking and gesture recognition system, with gaze and head movement detector by ICCS [7], with Pure Data [8], a graphical programming environment for real-time audio processing, and with a face detection module that can detect user's smiles, head nods and shakes [1].

LISTENER INTENT PLANNER. The **Listener Intent Planner** LIT computes the backchannels of the agent when being a listener in 2 steps, when to do a backchannel and which one. LIT takes as input audio-visual data and selects which user's behaviors (for example, a head nod or a variation of pitch of the user's voice) could elicit a backchannel from the agent [3]. Then LIT decides which communicative intentions (agreement, refusal, interest...)

the agent will convey. The communicative intentions are written in FML-APML language and sent to the Behavior Realizer module via Psyclone.

BEHAVIOR PLANNER. The Behavior Planner takes as input both the agent's communicative intentions specified in FML-APML and the agent's *baseline*. The agent’s baseline contains information on the preference the agent has in using modality (head, gaze, face, gesture, and torso) and on the expressive quality of each of these modalities. *Expressivity* is defined by a set of parameters that affect the qualities of the agent's behavior: e.g. wide vs. narrow gestures, smooth vs. jerky movements. Depending on the agent's baseline, the system computes the *dynamicline* that corresponds to the local behavior tendency, associated with each intention. The dynamicline, together with the current communicative intention is also used to select the multimodal behavior that best conveys the given intention: e.g., an agent that is always low aroused (baseline), in a joy state (communicative intention) could produce just a light smile, without moving the rest of its body; while a very expressive agent (baseline) with the same communicative intention could produce a combination of signals like smiling, stretching his arms, all at the same time.

BEHAVIOR REALIZER. This module generates the animation of Greta in MPEG4 standard [6]. The input to the module is specified in BML language. It contains a text to be spoken by Greta and/or a set of nonverbal signals to be displayed. The expressivity and the intensity of each signal are also specified. The Behavior Realizer synchronizes the behaviors across modalities. It solves also eventual conflicts between the signals that use the same modality. The speech is generated by an external TTS and the lips movements are added to the animation. Finally, both the audio file in *wav* format and the animation file are sent to the FAP-BAP Player via Psyclone.

FAP-BAP PLAYER. The FAP-BAP Player receives the animation (a sequence of FAP frames) generated by the Behavior Realizer and plays it by drawing the Greta's head and body in a graphic window. The animation is generated using OpenGL library.

SYNCHRONIZATION. The synchronization of all modules in the distributed environment is ensured by the **Central Clock** which broadcasts regularly timestamps through Psyclone.

APPLICATION. Our architecture is very modular. It allows us to plug and play different modules. Within the eNTERFACE summer workshop [5] the Listener Intent Planner component has been connected both to our FAP-BAP Player and to an Aibo robot. The backchannel signals generated by our system were displayed both by our Greta and by the robot. Since Aibo is a dog-like robot, an ad hoc backchannel lexicon had to be elaborated. For example, when the backchannel signal requested is a smile the dog wags its tail and its ears turning on bright lights on its head and on its back. The control of the behaviors of the agent and of the robot was done via the BML language. Figure 2 shows our agent and Aibo.

3. CONCLUSION

We have described an architecture for Greta, an expressive ECA. Our system can generate listener behavior while interacting with the user. The video (http://www.tsi.enst.fr/~niewiado/AAMAS_09/video1.avi) shows the interaction between the human speaker and Greta in the role of the listener. In the video one can see Greta sending backchannel signal in reaction to user behavior.

Figure 2: Greta and the Aibo dog.

4. ACKNOWLEDGMENTS

We are very grateful to André-Marie Pez, Etienne de Sevin and Ken Prepin for their help. This work has been funded by the STREP SEMAINE project IST-211486 (http://www.semaine-project.eu), the IP-CALLAS project IST-034800 (http://www.callas-newmedia.eu) and by the ANR-MyBlog-3D (https://picoforge.int-evry.fr/cgi-bin/twiki/view/Myblog3d/Web).

5. REFERENCES

[1] Baklouti, M., Couvet, S., Monacelli, E. 2008. Intelligent Camera Interface (ICI): A Challenging HMI for Disabled People. In Advances in Computer-Human Interaction.

[2] Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., Vilhjalmsson, H. 2008. Why Conversational Agents do what they do? Functional Representations for Generating Conversational Agent Behavior. The First Functional Markup Language Workshop. Estoril, Portugal.

[3] Maatman, R. M., Gratch, J., Marsella, S. 2005. Natural behavior of a listening agent. In: 5th International Conference on Interactive Virtual Agents. Kos, Greece.

[4] Morency, L., Sidner, C., Lee, C., Darrell, T. 2005. Contextual recognition of head gestures. In: 7th International Conference on Multimodal Interfaces, pages 18-24. ACM New York, USA.

[5] Moubayed, S. A., Baklouti, M., Chetouani, M., Dutoit, T., Mahdhaoui, A., Martin, J.-C., Ondas, S., Pelachaud, C., Urbain, J., Yilmaz, M. 2008. Generating Robot and Agent Backchannels during a Storytelling Experiment. In ICRA, Pasadena, California.

[6] Ostermann, J. 2002. Face Animation in MPEG-4, in: Pandzic, I.S., Forchheimer, R., (eds.), MPEG-4 Facial Animation - The Standard Implementation and Applications, Wiley, England, pages 17-55.

[7] Peters, C., Asteriadis, S., Karpouzis, K., de Sevin, E. 2008. Towards a real-time gaze-based shared attention for a virtual agent. In: Workshop on Affective Interaction in Natural Environments (AFFINE), Crete, Greece.

[8] Pure data. http://www.puredata.org.

[9] Thórisson, K.R., List, T., Pennock, C., DiPirro, J. 2005. Whiteboards: Scheduling blackboards for interactive robots. In Twentieth National Conference on Artificial Intelligence.

[10] Vilhjalmsson, H., Cantelmo, N., Cassell, J., Ech Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thörisson, K. R., van Welbergen, H., van der Werf, R. 2007. The Behavior Markup Language: Recent developments and challenges. In 7th International Conference on Intelligent Virtual Agents.