Decalog 2007: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pages 179–180.
Trento, Italy, 30 May – 1 June 2007. Edited by Ron Artstein and Laure Vieu.

AMBROSIO: THE MIMUS TALKING HEAD

Pilar Manchón, Antonio Ávila, David Ávila, Guillermo Pérez, Gabriel Amores
University of Seville
{pmanchon, aavila, davila, gperez, jgabriel}@us.es

ABSTRACT

In a technology-driven world where HCI is developing fast and the research on Multimodal Dialogue Systems is substantial, personality endowed virtual characters are acquiring importance. Socio-psychological research [4] indicates that users are in general more willing to interact with technology when the latter is 'humanized', that is, when the interaction is closer to that between humans. In this paper, the integration of a talking head in MIMUS, a home-control multimodal dialogue system, is presented. Also known as 'Ambrosio', the MIMUS talking head in synchronized with state-of-the-art synthesizers and provides a full range of facial expressions and motions. A preliminary evaluation of the talking head impact on the overall system perception seems to confirm the benefits of integrating virtual characters in this type of systems.

Index Terms—Multimodal *dialogue systems and* ***interfaces, HCI,*** *human factors, talking heads, personality,*

1. INTRODUCTION

During the last couple of decades at least, a great deal of research around Multimodal Dialogue Systems is being conducted. Among other interesting issues, the importance of introducing virtual entities to interact with users has been analysed from different perspectives.

Some authors advocate for the benefits of human-like interaction, endowing virtual characters with human characteristics to make human-machine interaction as close as possible as human-human interaction [4]. Other authors however are reluctant to believe that simulating human communication is the best alternative. Moreover, they express serious concerns about negative social and cognitive issues, pointing at a rather dramatic clash of the human-computer social order in cognitive terms [9].

In the development of the MIMUS system, we have opted for the human-centered design approach and implemented a talking head with the ability to talk, change its facial expressions and perform some motions. The overall purpose of the MIMUS system is to become a practical and valuable tool in the smart home scenario, and more in particular, an everyday tool for the specific focus group the data was gathered from in a set of WoZ experiments [10]. It is therefore understandable that with that objective in mind, different sub--objectives gain importance: human--like interaction must be not only efficient, but may and/or should also include additional human features. In order to endow the system with sufficient capabilities to fulfill these requirements, the MIMUS system has been furnished with the hereby described talking head that complements the system's personality, and confers an appearance of human—like communication on the interaction.

2. THE MIMUS SYSTEM

MIMUS is a multimodal dialogue system for the control of a smart home. It relies of a flexible architecture that allows for the integration of multiple input and output modalities. The current design and configuration responds to the requirements of the selected focus group of users (wheel-chair bound users), although there is no reason why it could not be reconfigured for different user profiles. As a matter of fact, one of the main advantages of the flexile architecture above mentioned is the possibility of configuring the system's behavior in terms of the information available at user profile level.

As in [5], MIMUS offers a pseudo-symmetric architecture: graphical and voice modalities are available both at input and output, although written text is provided as output but it is not a current input option. Nonetheless, MIMUS offers additional advantages since any functionality can be achieved by mixing modalities, using only voice, or only graphically. This is particularly important to allow for different user profiles to take full advantage of the system, especially those with special needs.

MIMUS consists of a set of collaborative OAA agents. It consists of an ISU-based dialogue manager [6], a knowledge manager, a device manager, ASR and TTS managers, several graphical agents and the talking head. This Talking Head complements previous implementations, and endows the system with additional features and communicative intensity.

The literature [4] illustrates how different experiments show that computers are indeed social actors, and also that the users' conduct is quite different when interacting with a virtual character as opposed to when they interact with a faceless computer. The overall user satisfaction is greater when interacting with a virtual character. MIMUS seeks to be a clear example of user--centered design, and with the user always in mind, the MIMUS talking head has been integrated into the main system architecture. For more information about the overall system architecture, please see [7].

3. AMBROSIO'S DESIGN AND IMPLEMENTATION

Endowing the character with a name has a manifold purpose: Personalization (users can give him a name of their choice),

Personification (they will address the system at personal level, reinforcing the sense of human--like communication) and Voice activation (Ambrosio will remain inactive until called for duty).

Ambrosio has been implemented in 3D to allow for more natural and realistic gestures and movements. The graphical engine used is OGRE, a powerful, free and easy to use tool.

3.1. 3D FACIAL ANIMATION

Modeling: The modeling methodology chosen is based on the facial muscular structure [2], which determines the basic modeling lines and areas that will in turn allow for the generation of facial expressions.
Expressiveness: In a 3D real--time application, facial expressions are generated by means of pre--defined poses. Once each expression is modeled, the 3D vertex variation for each pose is recorded separately, so that several expressions can be simultaneously generated. This is a widely used method called lineal interpolation animation.
Animation: It is achieved throughout a skeleton system: each bone has an impact on the neighboring vertexes. Each vertex has an associated list where each bone has an associated value ranging from 0 to 1.
Texture: For picture-like realism, the light has been integrated in the texture in order to achieve better performance and good graphical quality.

3.2. ARCHITECTURE & EXPRESIVENESS

As in [3], the system consists of four different subsystems: Input, Synchronization, Speech synthesis and Face management. The current talking head is integrated with Loquendo, which allows for lip synchronization.

According to the literature [1], gestures and expressions are almost or quite as important as speech itself: it reinforces the overall communicative act. It is therefore mandatory to determine the different behavior layers, and establish which gestures are conscious and which unconscious. Ambrosio's conscious motions are nodding and shaking; his unconscious motions are breathing and blinking. His expressions coincide mostly with the standard [1] (happiness, sadness, anger, fear and surprise). However, "disgust" has been substituted by "doubt", being the latter more useful in the dialogue context.

4. CONCLUSIONS & FUTURE WORK

As conclude in previous research [4] and our own experiments, a human-like talking head has a significant positive impact on the subjects' perception and willingness to use MIMUS. Although no formal evaluation of the system has taken place yet, MIMUS has already been presented successfully in different forums, and as expected, ``Ambrosio'' has always made quite an impression, making the system more appealing to use and approachable. In the opinion of the potential users enquired, Ambrosio's motions and expressions were helpful and communicative. Ambrosio's desing and architecture are the result of the integration of different theoretical and practical approaches to avatars, personality and dialogue, all applied to a smart home system. Its flexible architecture will allow for very interesting extensions in the future such as user-taylored personalities. The final objective is to generate a range of different users' profiles, detect the user's personality direct or indirectly, and establish their direct mapping with compatible virtual characters that will match the users' personalities and preferences.

Future developments necessarily entail a formal overall system evaluation, but also a specific human factors and usability evaluation of Ambrosio.

5. ACKNOWLEDGEMENTS

This work was carried out under the "TALK" research project, funded by EU´s FP6 [ref. 507802] and the "Generación de Lenguaje Natural para Sistemas de Diálogo" project, funded by the Spanish Ministry of Education under grant TIN2006-14433-C02-02.

6. REFERENCES

[1] Fischer, A, Ostermann J., Beutnagel M. and Wang, Y. Integration of talking heads and text.to.speech synthesizers fot visual tts. In International Conference of Speech and Language Processing, Sydney, Australia, 1998. ICSLP98.

[2] Moreaux, A. Anatomie Artistique de l'Homme. Maloine, Paris, 2nd edition, 1990.

[3] Gattass, M. .Lucena P. S. and Velho L.. Expressive talking heads: A study on speech and facial expression in virtual characters. Scientia, 13(2):1.12, 2002.

[4] Reeves, B. and Nass. C. The Media Equation. CSLI.Cambridge University Press, 1996.

[5] Wahlster, W. "SmartKom: Symmetric Multimodality in an Adaptive and Reusable Dialogue Shell", in Proceedings of the HCI Status Conference, Berlin, Germany, pp. 47-62., 2003

[6] Traum, D. Bos, J., Cooper, R. Larsson, S. Lewin, I. Matheson, C. and Poesio, M.. A model of Dialogue Moves and Information State Revision. Technical Report D2.1, Trindi Project. 1999.

[7] Pérez, G., Amores G. and Manchón, P.. A Multimodal Architecture for Home Control by Disabled Users. (2006) Proceedings of IEEE/ACL Workshop on Spoken Language Technology (SLT), Aruba. December 2006.

[8] P. Manchón, D. Ávila & A. Ávila. Modality-specific Resources: Extension. Deliverable 3.3 Extension. TALK Project. December 2006.

[9] Schmidt, C. T. A. (2005). Of Robots and Believing. Minds and Machines. Kluwer Academic Publishers, Hingham, MA, USA, 15: 195-205.

[10] P. Manchón, C. del Solar, G. Amores & G. Pérez (2006) "The MIMUS Corpus." LREC 2006 International Workshop on Multimodal Corpora From Multimodal Behaviour Theories to Usable Models. Genoa, Italy, pp. 56-59