

EMBODIED CONVERSATIONAL AGENTS IN VERBAL AND NON-VERBAL COMMUNICATION

Björn Granström

KTH - Royal Institute of Technology, Sweden

RESUMEN

In face-to-face communication both visual and auditory information play an obvious and significant role. Traditionally in phonetic research the auditory effects of speech production have been the primary object of study. However, when it comes to the non-verbal aspects of speech communication the primary nature of acoustics is not as evident, and understanding the interactions between visual expressions, dialogue functions and the acoustics of the corresponding speech presents a substantial challenge. Some of the visual articulation is for obvious reasons closely related to the speech acoustics (e.g. movements of the lips and jaw), but there is other articulatory movement affecting speech acoustics that is not visible on the outside of the face. On the other hand, many facial gestures used for communicative purposes do not affect the acoustics directly, but might nevertheless be connected on a higher communicative level in which the timing of the gestures could play an important role. The context of much of our research regarding these questions is to be able to create an animated talking agent capable of displaying realistic communicative behavior and suitable for use in conversational spoken language systems.

Useful applications of talking heads include aids for the hearing impaired, educational software, audiovisual human perception experiments, entertainment, and high quality audiovisual text-to-speech synthesis for applications such as news reading. The use of the talking head aims at increasing effectiveness by building on the user's social skills to improve the flow of the dialogue. Visual cues to feedback, turntaking and signaling the system's internal state are key aspects of effective interaction.

The focus of this paper is to present an overview of some of the research involved in the development of audiovisual synthesis to improve the talking head. Some examples of results and applications involving the analysis and modeling of acoustic and visual aspects of verbal and non-verbal communication are presented.

Granström, B., & House, D. (2007). Inside out - Acoustic and visual aspects of verbal and non-verbal communication (Keynote Paper). Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 11-18.

Granström, B., & House, D. (2007). Modelling and evaluating verbal and non-verbal communication in talking animated interface agents. In Dybkjaer, I., Hemsén, H., & Minker, W. (Eds.), *Evaluation of Text and Speech Systems* (pp. 65-98). Springer-Verlag Ltd.

Beskow, J., Granström, B., & House, D. (2007). Analysis and synthesis of multimodal verbal and non-verbal interaction for animated interface agents. In Esposito, A., Faundez-Zanuy, M., Keller, E., & Marinaro, M. (Eds.), *Verbal and Nonverbal Communication Behaviours* (pp. 250-263). Berlin: Springer-Verlag.

Björn Granström joined KTH, The Royal Institute of Technology, in Stockholm, Sweden in 1969, after graduating as MSc in Electrical Engineering. After further studies in Phonetics and General Linguistics at Stockholm University he became Doctor of Science at KTH in 1977 with the thesis "Perception and Synthesis of Speech". In 1987 he replaced Gunnar Fant as Professor in Speech Communication. He has been the director of CTT, The Center for Speech Technology <<http://www.speech.kth.se/ctt>>, since its start in 1996. Granström has numerous publications in the speech research and technology area. The group is actively participating in many EU projects. His recent research interests include multi-modal communication systems using embodied conversational agents.