Timing and Rhythm in Multimodal Communication for Conversational Agents

Ipke Wachsmuth (ipke@techfak.uni-bielefeld.de)
Faculty of Technology, University of Bielefeld
D-33594 Bielefeld, Germany

Motivation

Synthesis of lifelike gesture is finding growing attention in human-computer interaction. In particular, synchronization of synthetic gestures with speech output is one of the goals for embodied conversational agents which have become a new paradigm for the study of gesture and for human-computer interface (Cassell et al., 2000). Embodied conversational agents are computer-generated characters that resemble similar properties as humans in verbal and nonverbal face-to-face conversation.

Gesture production in humans is a complex process leading to characteristic shape and dynamic properties of gestures which enable humans to distinguish them from subsidiary movement and recognize them as meaningful. In coverbal gestures the stroke (the most effortful part of the gesture) is tightly coupled to accompanying speech, yielding semantic, pragmatic, and temporal synchrony between the two modalities (McNeill, 1992).

Although promising work exists for the production of synthetic gestures, natural timing for the gesture stroke and synchronizing it with speech output remains a research challenge. For instance, the REA system by Cassell and coworkers (in Cassell et al., 2000) implements an embodied agent that produces verbal and gestural output. Yet though precise timing of spoken and gestural utterances is targetted in their work, the authors state that it has not been satisfactorily solved.

Articulated Communicator

A mid-range goal of our research is the conception of an “articulated communicator” that conducts multimodal dialog with a human partner in cooperating on a model airplane construction task. In this context an operational model was developed that enables lifelike gesture animations to be rendered in real time from representations of spatiotemporal gesture knowledge (Kopp & Wachsmuth, 2000). Based on various findings on the production of human gesture, the model provides means for motion representation, planning, and control to drive the kinematic skeleton of a figure which comprises 43 degrees of freedom in 29 joints for the main body and 20 DOF for each hand (see Figure 1). A movement plan is formed as a tree representation of a temporally ordered set of movement constraints in three steps:

(1) retrieve feature-based specification from a gestuary
(2) adapt it to the individual gesture context
(3) qualify temporal movement constraints in accordance with external timing constraints.

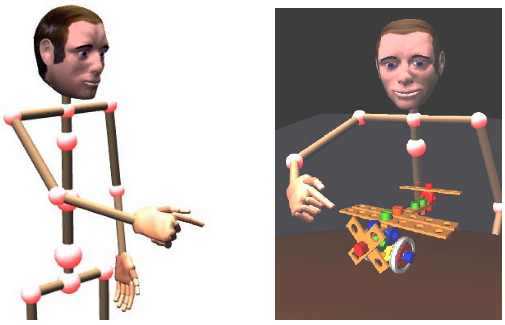

Figure 1: Articulated Communicator.

Hence our model is conceived to enable cross-modal synchrony with respect to the coordination of gestures with the signal generated by a text-to-speech system. In multimodal communication, by which we mean the concurrent formation of utterances that include gesture and speech, a rhythmic alternation of phases of tension and relaxation can be observed. The issue of rhythm in communication has been addressed widely and has been a key idea in our earlier work on synchronizing gesture and speech in HCI input devices (Wachsmuth, 1999). Achieving precise timing for accented parts in the gesture stroke as a basis to synchronize them with stressed syllables in speech is work currently in progress.

Acknowledgment

This research is partially supported by the Deutsche Forschungsgemeinschaft in the Collaborative Research Center “Situated Artificial Communicators” (SFB 360).

References

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.) (2000). *Embodied Conversational Agents*. Cambridge (MA): The MIT Press.

Kopp, S., & Wachsmuth, I. (2000). A knowledge-based approach for lifelike gesture animation. *ECAI 2000 Proc. 14th European Conf. on Artificial Intelligence*. Amsterdam: IOS Press.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Wachsmuth, I. (1999). Communicative Rhythm in Gesture and Speech. In A. Braffort et al. (Eds.), *Gesture-based Communication in Human-Computer Interaction.* Berlin: Springer (LNAI 1739).