

Improving human-system communication with talking heads

David House and Björn Granström

Centre for Speech Technology (CTT), Dept. of Speech, Music and Hearing, KTH, Stockholm, Sweden

At CTT we have for the past several years been developing spoken dialogue applications that include animated talking agents. Our motivation for moving into audiovisual output is to investigate the advantages of multimodality in human-system communication. While the mainstream character animation area has focussed on the naturalness and realism of the animated agents, our primary concern has been the possible increase of intelligibility and efficiency of interaction, resulting from the addition of a talking face. In our first dialogue system, Waxholm, the agent used the deictic function of indicating specific information on the screen by eye gaze. In another project, Teleface, we were specifically concerned with the advantages in intelligibility that a talking face could provide. While the results showed that the intelligibility of the audiovisual speech synthesis was not as high as that of natural speech in most cases, there was a significant increase in intelligibility when compared to speech synthesis without the visual component.

In recent studies we have investigated the use of facial gesture cues to convey such dialogue-related functions as feedback and turntaking signals as well as prosodic functions such as prominence signalling. Results show that cues such as eyebrow and head movement can independently signal prominence. Current results also indicate that there can be considerable differences in cue strengths among visual cues such as smiling and nodding and that such cues can contribute in an additive manner together with auditory prosody as cues to dialogue functions. Results from some of these studies will be presented at the workshop along with examples of talking heads in spoken dialogue applications.