

Engagement and Floor Control in Hybrid Meetings

BY THE HUMAN MEDIA INTERACTION GROUP OF THE UNIVERSITY OF TWENTE

It is a frequently observed fact that remote participants in hybrid meetings often have problems to follow what is going on in the (physical) meeting room they are connected with. From extensive analyses of face to face and remote meetings it is clear how important non-verbal social behavior is in communicating who is being addressed and who is expected to take turn. One of the possible applications of AMIDA research in real-time automatic scene analyses and meeting behavior recognition is in the development of technology and interfaces that support participants in distributed meeting environment.

The Human Media Interaction group of the University of Twente has developed a User Engagement and Floor Control Demonstrator, a system that uses modules for online speech recognition, real-time visual focus of attention as well as a module that signals who is being addressed by the speaker. A built-in keyword spotter allows an automatic meeting assistant to call the remote participant's attention when a topic of interest is raised, pointing at the transcription of the fragment to help him catch-up. The first version of the UEFC demo was presented at the AMIDA review meeting in Edinburg last year.

In the final year of AMIDA we focussed on two main tasks. The first task is the integration of the UEFC demo and the Content Linking Demo. This has resulted in an offline version of the UEFC demo that demonstrates how both demonstrators can work with the same instance of the HUB database that contains the annotation layers of the meeting as well as the documents that are automatically retrieved based on a set of key phrases selected by the user.

Both the offline and the online version of the UEFC demo make use of the HMI Media Streaming package that handles the synchronisation, compression and streaming of video and audio for communication, processing and recording. The HMI Media Streaming software has also been used in the remote meeting experiments that were performed in an AMIDA Miniproject with TXchange, a member of the COI (see a previous issue of this Newsletter). The HMI media streaming package, based on DirectShow for MS Windows, allows easy development of user interfaces in a modular way.

The second and main task that has been performed is to experiment with two different user interface for videoconferencing: a classical 2D interface and a 3D interface. The experiments were performed in the meeting room of the new SmartXP Lab at the University of Twente. The main differences between both experimental conditions is that in the 3D version non-verbal communicative behavior in the form of gaze is transmitted in a substantially improved way. We compare a conventional video conferencing interface versus an interface where video streams were presented to remote participants in an integrated 3D environment, in a context where a small group of three



Figure: showing the 3D interface for the remote participant

co-located persons had a meeting joined by one person located on a remote place. The conventional interface we used was in essence like a Skype or Adobe connect interface in the sense that meeting participants are visible in separate video frames, and each of them would be looking straight into their own webcam.

In both conditions the video image of the remote participant was visible to the co-located persons on a classical, but large, video screen. Such multimodal interfaces offer already a lot in that both speech as well as facial expressions are communicated. But certain non-verbal behavioral aspects are still lacking, in particular body pose and gaze direction from one person to another. Our integrated interface tries to improve these aspects, aiming at enhancing the presence of participants. On the remote participant side it employs a basic 3D interface where the (video images of) other participants appear to be sitting around a virtual table, consistent with the real, physical situation (see Figure: showing the 3D interface for the remote participant). Cameras capturing each of the co-located participants were no longer in front of them, but rather repositioned, so that looking in the direction of the remote participant (screen) would coincide with looking into the camera, as far as possible.

The main intended effect of the 3D interface is that it becomes much easier to observe who is looking to whom. We arranged an experiment where ten groups, each consisting of four participants, held short meetings using the two different interfaces, while being observed by means of cameras as well as through a two-way mirror, allowing observers to track the participants gaze behavior, in order to measure the amount of attention that participants receive from others, and to analyze turn taking behavior. We also measured perceived presence by means of a social presence questionnaire, asking about aspects like (perceived) co-presence, message understanding, and attention allocation. Finally, we held interviews immediately after meeting sessions. The first results show interesting and statistically significant differences between the two experimental situations, mostly favoring our integrated interface. Ambiguity with respect to who is looking to whom in the classical interface has been observed, and might at least partially explain these results. We expect that the corpus of synchronized audio and video recordings, together with the questionnaires are a valuable data set for future research in remote meeting behavior.

Rieks op den Akker, Job Zwiars, Hendri Hondorp, Betsy van Dijk, Olga Kulyk, Dennis Hof, Anton Nijholt, Dennis Reidsma,

*Human Media Interaction, University of Twente,
Enschede the Netherlands
infrieks@cs.utwente.nl*