

# Conversational Awareness in Multiparty VMC

Roel Vertegaal

Dept. of Ergonomics, University of Twente  
P.O. Box 217, 7500 AE Enschede, The Netherlands  
*roel@acm.org*

## ABSTRACT

In this demonstration, we present a number of videoconferencing systems which differ in support for conversational awareness. We argue that such systems should convey speech, relative position, gaze direction and gaze of the participants, but not necessarily full-motion video.

## Keywords

CSCW, groupware, videoconferencing, awareness, attention.

## INTRODUCTION

Face-to-face communication is generally regarded as an optimal form of synchronous interactive human communication. With increased bandwidth availability, efforts towards optimizing telecommunication are therefore largely aimed at closer modeling of face-to-face situations. It was thought that this could be achieved by adding video images to existing telephony-based applications. Such Video Mediated Communication (VMC) systems would increase the available sensory bandwidth, thus making mediated communication more like face-to-face communication. The visual channel in VMC systems would support the use of visual cues and this information richness would stimulate a sense of social presence amongst the remote conversational partners. A typical VMC system is shown in figure 1. Users see themselves and the other participants arranged in a grid on their screen. They can also speak to each other. Images are conveyed by a camera at top of their screen.



Fig 1. A typical VMC setup.

VMC has been regarded as a superior alternative to audio-only communication systems and a substitute for face-to-face communication. However, some organizations planning to adopt VMC as a core communications strategy are reconsidering their options. According to Van der Velden, VMC systems such as the one shown in fig. 1 are not well suited for exchanging information during complex tasks or for activities in which relations between conversational partners play an important role [5].

Empirical studies [3, 4, 5, 6] to both the short- and long-term effects of different mediated systems on group tasks seem to point in the following directions:

1) In terms of performance and other measures, differences between groups working together by means of an audio-only communication link and groups working together by means of VMC appear to be relatively small.

2) To what extent conversational partners think they can affect each other's behaviour seems to be inherently limited by the fact that their communication is mediated. This makes it unlikely that VMC will ever be a complete substitute for face-to-face communication, however realistically it models it.

3) Although a VMC system can convey cues additional to speech such as lip movements, facial expressions and gestures, its use for establishing context-awareness amongst participants could be of greater value to group functioning, particularly when many participants are involved.

Recent research efforts aim at improving VMC by modeling a face-to-face situation as realistically as possible. In the MAJIC system [2], this is done by using life-size images and by integrating these images in participants' real work environments. Such efforts go at a considerable cost in terms of network and other resources. We feel it is only those provisions in the MAJIC system that support awareness (by conveying the participants' attention) that make it a better tool for group communication. We question the necessity or even the desirability of providing face-to-face realism in VMC systems. If VMC is not like face-to-face communication, our attempts to make it look that way may in fact limit the user's understanding of how to use it effectively. Instead, we should aim at using the visual channel to support different levels of awareness by supplementing speech communication with attention-related data at different levels of refinement. Such systems would aim at modeling those characteristics of face-to-face meetings which are crucial to understanding who is talking to whom about what. Whether video images are life-size or even whether they are full-motion seems to be of secondary importance in terms of usability considerations and may simply be dependent on network limitations.

## CONVERSATIONAL AWARENESS

Given that VMC is to be used to supplement rather than substitute face-to-face communication, and that for small groups audio-only communication may serve equally well, the main problem in supporting multiparty conferences is to provide awareness about whom other participants are addressing, and what others are doing. On a macro level of awareness, knowing what others are doing and whether they are available for communication is important. One only

needs thumbnail low-frequency images to convey this. On a micro level of awareness, knowing who is talking to whom (conversational awareness), and who is working on what (workspace awareness) are important. Most commercial VMC systems are limited in their support for this micro level of awareness since they do not preserve spatial features of auditory and visual cues. For VMC systems to support conversational awareness we define the following incremental requirements:

1) Relative position: relative viewpoints of the participants should be based on a common reference point (e.g., around a shared workspace), providing basic support for deictic references. A corresponding spatial separation of the audio sources (e.g. by means of stereo panning) eases selective listening, for example during side conversations.

2) Gaze direction: the absolute direction in which someone looks should be conveyed. It allows deictic references (e.g., 'look at this') and allows participants to determine who is speaking to whom, important in turn-taking and constituting side conversations.

3) Gaze: A special case of gaze direction. Allowing participants to gaze at each other's facial region eases turn-taking: speakers can see whether listeners are still attending, and can use gaze to keep the floor when pausing [1]. Mutual gaze constitutes eye contact.

#### Supporting Conversational Awareness in VMC

We have experimented with different hardware setups, each with different support for conversational awareness:

System 1. Traditional VMC setup: Full-motion video without relative positioning (as in Fig. 1).

System 2. Full-motion video with relative positioning: each participant has only one camera, but participants are arranged on a screen according to a round table setting, possibly around a shared on-screen workspace (Fig. 2).

System 3. Full-motion video with relative positioning conveying gaze direction but not gaze: each participant is represented by a camera-screen unit, as with the Hydra system [4]. Units are placed in an angular setting, some distance apart (Fig. 3).

System 4. Full-motion video with relative positioning conveying both gaze direction and gaze: each camera-screen unit is a videotunnel. By means of a half-silvered mirror, gaze is simulated by allowing participants to look into the camera when they look at a screen (as in Fig. 3, but with a video tunnel for each screen).

System 5. Still video with relative positioning, gaze direction and gaze. This is a special case in which each participant is shown pre-recorded still images of the other participants. Each still image shows a participant gazing in a particular direction (Fig. 4). Gaze direction and gaze are simulated by selectively displaying these images according to where the participants look. This can be done by registering each participant's eye movements using an eye-tracking system. The setup is similar to Fig. 2, with an eye-tracking system replacing the camera. Only audio connections are needed, supplemented with low-bandwidth eye co-ordinate data.

During our demonstration, attendees will be able to use some of the above systems to experience the impact of conversational awareness on the communication process.

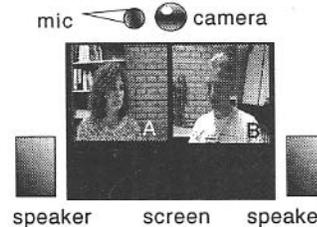


Fig 2. VMC with relative positioning only.

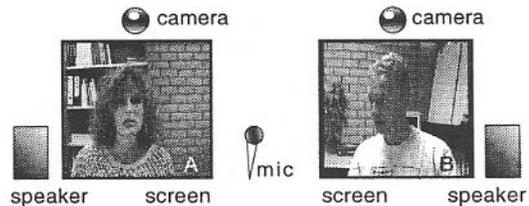


Fig 3. VMC with relative positioning and gaze direction.



Fig 4. Still images at 0, 30 and 60 degree angles.

#### CONCLUSIONS

To date, the differences found in terms of task performance and other measures between groups using the different described VMC systems have been relatively small. In part, this may be due to the small size of the groups used in most studies. However, it seems likely that for small, well-structured groups, audio-only communication supplemented with conversational awareness data (System 5) should offer task performance similar to that of the other described systems, with much lower network bandwidth requirements. In unstructured groups larger than four, it seems likely that systems which provide conversational awareness will in fact increase task performance. We are currently conducting experiments to seek empirical support for this claim.

#### REFERENCES

1. Argyle, M. *The Psychology of Interpersonal Behaviour*. London: Penguin Books, 1967.
2. Okada, K.-I., Maeda, F., Ichikawa, Y. and Matsushita, Y. Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. *Proceedings of ACM CSCW'94*, 1994.
3. Rutter, D.R. and Stephenson, G.M. The Role of Visual Communication in Synchronising Conversation. *European Journal of Social Psychology* 7, 1977.
4. Sellen, A.J. Speech Patterns in Video-Mediated Conversations. *Proceedings of ACM CHI'92*, 1992.
5. Velden, J. v.d. *Samenwerken op afstand*. PhD Thesis. Technical University of Delft, The Netherlands, 1995.
6. Vons, J.A. *Human Interaction through Video Mediated Systems*. Master Thesis. Dept. of Ergonomics, University of Twente, The Netherlands, 1996.