Evaluating Advanced Video-Conferencing Interfaces for Remote Teamwork

1.   INTRODUCTION

Video-mediated communication (VMC) offers groups of geographically dispersed people the possibility of rich real-time communication and collaboration. Being able to see each others' faces while speaking to one another is generally assumed to be beneficial and desirable, allowing remote participants to communicate, feel, and interact with each other in approximately the same manner as they would do if they were actually together in the same room. Yet, despite the obvious advantages of visual communication over audio-only conversations, VMC still feels distant, artificial, cumbersome, and detached compared to being face-to-face.
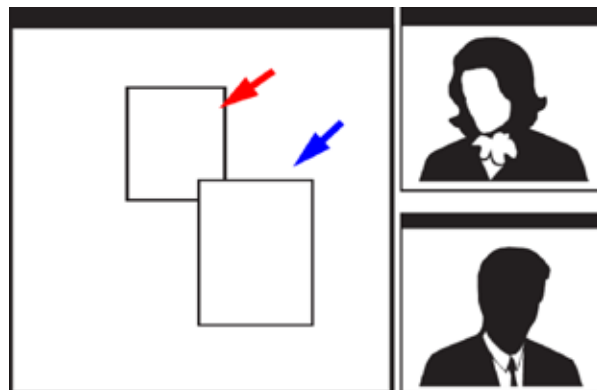

Figure 1: A standard video-conferencing layout: the video images of two participants are displayed alongside a shared workspace which they are jointly interacting with.

One shortcoming of common video-conferencing systems (see Figure 1) which contributes to this feeling is that the 3D context between people and their shared workspace is lost. It is therefore not possible for participants to tell from the video of others what they are looking at, what they are working on, or who they are talking to – all of which can cause issues for coordinating their collaborative activities.

Researchers in the field of CSCW (computer-supported cooperative work) have tried to improve VMC through the development of innovative video-conferencing interfaces which support more natural, easy, and efficient video-collaboration.[1]

In this paper I describe my attempt to design and evaluate video-conferencing prototypes that support several spatial aspects of face-to-face meetings which are otherwise absent in regular video-conferencing systems. In line with a user-centered design philosophy, I believe that thorough evaluation is a pivotal part of an iterative design process working towards more usable, more efficient and more fun-to-use video-conferencing systems.

In section 2, I give an overview of possible evaluation criteria that assess the quality of a video-conferencing interface. Section 3 then presents the design and the findings of a user study that applied some of the previously introduced measures. Finally, section 4 discusses the findings and concludes this paper.

2.   EVALUATING VIDEO-CONFERENCING INTERFACES

The goal of any real-time telecommunication media is to collapse the space between geographically dispersed groups and create the illusion that people are together, when in fact they are not. Or, in other words, the more a communications medium supports face-to-face-like communication and social interaction, the better.

Based on this premise, the quality of different video-conferencing systems can be assessed with respect to the degree to which they allow remote users to feel and interact with each other as if they were actually together in the same location.

In this regard, the concept of "social presence" provides a framework for understanding and measuring the differences between mediated and non-mediated communication. Short et al initially defined social presence as an experienced characteristic of a communications medium that depends on the "salience of the other person in the interaction."[2] Since then, the definition has been broadened to include the "feeling that the people with whom one is collaborating are in the same room,"[3] the "perceptual illusion of non-mediation",[4] or the "feeling that one has some level of access or insight into the other's intentional, cognitive, or affective states."[5] At a basic level, and as common denominator of all definitions of the term, social presence describes and encompasses different facets of the "sense of being together."[6]

Generally, social presence increases with the amount of supported nonverbal communication cues, the possibilities for immediate feedback, and the degree to which they support interpersonal rather than mere factual aspects in conversations.

In decreasing order of social presence, face-to-face communication is rated highest, followed by visual media such as video-conferencing, non-visual media such as the telephone, and

written media such as a business letter. However, two versions of the same medium can also support different levels of social presence if they support different forms of nonverbal communication.

Social presence is a quality of a medium that cannot be directly observed, but rather has to be measured indirectly. Generally, subjective measures and behavioural measures are distinguished.

Subjective social presence measures assess the "sense of being together" by asking questions of participants after they have used a telecommunication system to connect to a remote person. Questions are applied in a post-experience questionnaire or in interviews and can be directed towards the communication medium (see Table 1 for examples) or may directly target the experience of the other person (see Table 2 for examples). Telecommunication media which support a higher level of social presence are normally rated as warm, personal, and sociable.[7]

I RATE THE TYPE OF MEDIUM I JUST USED TO COLLABORATE AS:

Cold	$1-2-3-4-5-6-7$	Warm
Impersonal	$1-2-3-4-5-6-7$	Personal
Unsociable	$1-2-3-4-5-6-7$	Sociable

Table 1: Examples of items assessing social presence through ratings of the medium.

PLEASE READ THE STATEMENTS BELOW AND INDICATE YOUR DEGREE OF AGREEMENT WITH EACH STATEMENT.

It felt as if my partner and I were in the same room.

| Strongly disagree | $1-2-3-4-5-6-7$ | Strongly agree |

It was just like being face-to-face with my partner.

| Strongly disagree | $1-2-3-4-5-6-7$ | Strongly agree |

I was always aware of my partner's presence.

| Strongly disagree | $1-2-3-4-5-6-7$ | Strongly agree |

Table 2: Examples of items assessing social presence through ratings of the collaborative experience.

Questionnaires are relatively easy to use and inexpensive to administer. However, they may produce unstable and inconsistent user responses if subjects interpret questions differently, or if subjects misjudge their own telecommunication experience. It is therefore recommended to combine subjective social presence measures with objective behavioural measures.

Behavioural social presence measures are based on social interactions between remote interlocutors which occur automatically, without conscious thought. For example, the conversations of two persons who are collaborating through a video-conferencing system can be recorded and analysed for content and format characteristics. Mediated conversations differ from face-to-face communications in a number of ways. When using telecommunication media, people talk longer in one turn, interrupt each other less, or talk over one another less often. Also, in the absence of some non-verbal communication cues, people coordinate their actions and clarify issues in a more verbal form.

Telecommunication systems support a higher level of social presence if people talk and interact with each other in a way that is similar to their behaviour if they were in the same room.

The following section presents a study of social presence in different video-based telecommunication systems. The goal of this study was to gain a better understanding of how the design of the human-computer interface influences a sense of social presence which will in turn guide the improvement of similar systems in the future.

3.   EXPERIMENT – "SPATIALITY IN VIDEO-CONFERENCING"

The user experiment presented in this chapter investigated whether adding a sense of spatiality in VMC increases social presence. A more detailed description of the design and results of this study can be found in Hauber et al.[8]

Spatiality concerns an interface's level of support for fundamental physical spatial properties such as containment, distance, orientation, and movement.[9] Regular video-conferencing interfaces (such as the one in Figure 1) provide little spatiality. This is because participants' videos and their shared workspace are spatially decoupled and thus the spatial context between them is lost. The spatiality of a video-conferencing interface can be increased by re-creating a shared three-dimensional reference frame into which the shared workspace and the videos of participants are integrated in a spatially consistent way.

In this experiment, two approaches for adding spatial cues were compared within a standard video-conferencing interface and a face-to-face "gold standard" control. One spatial approach followed the concept of Video Collaborative Virtual Environment (video-CVEs), while the other was based on video streams in a physically fixed arrangement around an interactive table (both explained in more detail below).

The question at stake was whether the additional non-verbal communication cues in the spatial interfaces (e.g., one participant could tell where the other was looking) would result in higher social presence, measured both by subjective responses of participants and analysis of the recorded conversation.

3.1 Experiment Design

The scenario and collaborative task for this experiment were chosen based on a common face-to-face collaborative situation, where aspects of spatiality play an important role: people who are gathered around a table to discuss a set of photos.

- *task*

A photo-based collaborative task was designed for teams of two participants to work on. The task required them to work together matching photographs of dogs to photographs of their owners, entirely based on casual reports that people resemble their pets. The task was deliberately chosen to be "ill-defined," that is, to involve a lot of uncertainty resolution, because these types of tasks require rich communication between collaborators to come up with a solution that both agree on. Photographs of dogs and their owners were taken at a local beach with the consent of all owners.

- *participants*

Fifteen teams of two same-gender friends (eleven male teams, four female teams) were recruited among staff and students at the University of Canterbury. Each team worked on four rounds of the dog-matching task under the following four conditions.

- *experiment conditions*

1. Condition "Face-to-Face" (FtF): Unmediated face-to-face collaboration around printed photographs placed on a real table (see Figure 2a).

2. Condition "Video-Table" (vTAB): Mediated remote collaboration around a shared interactive table (see Figure 2b). Spatiality aspects were supported within a local, real-world reference frame. The digital photographs were displayed and prearranged on a touch-sensitive table surface that allowed for interaction with the pictures.

3. Condition "Standard video-conferencing" (sVC): Mediated collaboration through a standard video-conferencing interface (see Figure 2c). No aspects of a shared three-dimensional reference frame were given. This set-up used a state-of-the-art video-conferencing system (Conference XP) involving video streams of both participants displayed on the screen as well as a shared application window.

4. Condition "video-CVE" (vCVE): Mediated collaboration around a virtual table in a video-CVE (see Figure 2d). Spatiality aspects were supported within the remote, virtual space. While the interaction with digital photographs was done with a standard computer mouse, the representations of the table and of the participants' video streams were shown in the simulated three-dimensional space. Head-tracking data was used to control the individual view into the virtual room. That way, person A could (for example) change his or her viewpoint between the table and the video avatar of person B by moving his or her head up and down. At the same time, the orientation of person A's video avatar consistently followed their head movements, allowing person B in turn to infer what was in the view field, and thus the point of attention, of person A. The interface was based on the cAR/PE! virtual tele-collaboration space.[10]

a) Condition FtF

b) Condition vTAB

c) Condition sVC

d) Condition vCVE

Figure 2: Conditions in experiment "spatiality in video-conferencing" (©ACM).

- *procedure*

For every one-hour session a group of two subjects was present. Participants were given an overview of the study and first filled out a questionnaire capturing basic demographic information. Then, each team took part in four rounds, one round for each condition (FtF, vTAB, sVC, vCVE). The order of conditions was controlled beforehand following a Latin Square scheme.

The systems were explained to participants before each round and participants had the chance to make themselves familiar with the controls in a short warm-up round in each condition.

The task in each condition was the same. However, new sets of photographs with different dogs and owners were used each time. Each round finished as soon as participants signalled that they had found a combination of dogs and owners which they both agreed with.

After each round, they were asked to fill out an experiment questionnaire. After the fourth and final round was over, and the fourth questionnaire was filled out by the participants, they were briefly interviewed about how they liked the task and were asked to give their personal preference rankings out of all four conditions.

- *measures*

The experiment questionnaire included items such as those shown in Table 1 and 2 to assess social presence. Furthermore, videos of the participants were captured during each round of the experiment. Questionnaire data and the recorded videos were statistically analysed for significant differences between conditions after the experiment.

3.2 Results

All participants except one self-described "cat person" liked the task and quickly became engaged in finding the matching pairs. The most common judgment criteria were whether a dog would be a woman's or a man's dog, if a dog would match the more active or passive lifestyle inferred from the photographs of the owners, and matching hair colour and facial features between owners and dogs.

The teams' strategy of handling the photograph orientation was consistent in the face-to-face condition and the two spatial video-conferencing conditions. The two main strategies were to either rotate all pictures to be correctly oriented for person A first, and then rotate them all back so person B could have a look; or, to place the photographs in the middle of the table and rotate them about 90 degrees into a more neutral sideways position where both could examine them sideways at the same time.

Both spatial video-conferencing conditions yielded a marginally higher sense of social presence than the standard interface (see Figure 3). Face-to-face was confirmed to be the gold standard in social presence.
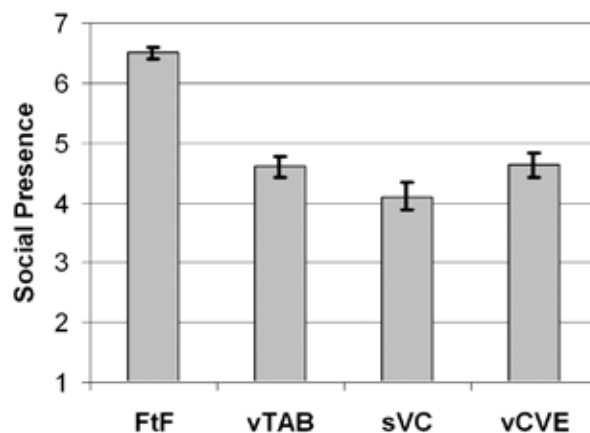

Figure 3: Averages and standard error of the degree of social presence based on collected questionnaire data.

The video analysis, however, revealed that collaboration in the standard video-conferencing interface was more efficient than the collaboration in the spatial video-conferencing conditions, while face-to-face collaboration was, as expected, the most efficient.

In the spatial video-conferencing conditions, participants took longer to come up with a solution. They also dedicated more of their spoken turns to coordinate their actions or to clarify issues.

Interaction in spatial interfaces was more complex due to the need for handling each participant's individual perspective of the shared workspace. For example, the photographs on the table had to be rotated repeatedly during the collaboration since they could only face one or the other user at a time. Interacting with digital photographs, however, is not as easy as interacting with paper prints that can be manipulated simultaneously by both participants using two-handed interaction. The additional mental effort introduced by handling and manipulating the photographs led to confusion and distraction from the task.

4.   DISCUSSION AND CONCLUSION

The spatial video-conferencing interfaces supported a higher level of social presence. This finding supported our main hypothesis: that adding a sense of spatiality in video-conferencing increases social presence.

However, when comparing collaboration in the spatial video-conferencing conditions and face-to-face collaboration, it became obvious that we had overvalued the role of seeing the other's face during the interaction. Collocated participants hardly ever looked at each other. Instead, the centre of attention during the collaboration was the shared table. While looking at the table, participants communicated and coordinated their activities by pointing, rotating, moving, ordering, aligning, and exchanging the photographs.

These subtle but pivotal elements of collaboration were supported to a much lesser degree in our spatial video-conferencing conditions, which led to the most salient differences in communication patterns. Participants were, for example, not able to manipulate the photographs simultaneously or use both hands for interaction, which inhibited efficient forms of communication.

The next steps in my research will therefore be to explore ways in which spatial video-conferencing interfaces can be further improved by providing novel interfaces that allow tangible, two-handed manipulation of shared objects.

Furthermore, I will continue to investigate other contributing factors that shape a sense of social presence in video-mediated communication. I am particularly interested in the impact of different types of collaborative tasks, but also in interactions between social presence and demographic variables such as gender, previous media experience, and personality traits.

ACKNOWLEDGEMENTS

1   For example, S Acker and S Levitt, "Designing Videoconferencing Facilities for Improved Eye Contact," *Journal of Broadcasting & Electronic Media*, 31 (1987), 181–91; H Ishii, M Kobayashi and K Arita, "Iterative Design of Seamless Collaboration Media," *Commun. ACM*, 37:8 (1994), 83–97; D Nguyen and J Canny, "Multiview: Spatially Faithful Group Video Conferencing," in *CHI '05: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York: ACM Press, 2005), 799–808.

2   J Short, E Williams, and B Christie, *The Social Psychology of Telecommunications* (New York: Wiley, 1976).

3   R Mason, *Using Communications Media in Open and Flexible Learning* (London: Kogan Page, 1994).

4   M Lombard and T Ditton, "At the Heart of it All: The Concept of Presence," *Journal of Computer Mediated Communication*, 3:2 (1997).

5   F Biocca and K Nowak, "Plugging your Body into the Telecommunication System: Mediated Embodiment, Media Interfaces, and Social Virtual Environments," in *Communication Technology and Society*, ed. C Lin and D Atkin (Hillsdale, NJ: Lawrence Erlbaum Associates, 2001), 57-124.

6   P de Greef and W A Ijsselsteijn, "Social Presence in a Home Teleapplication," *Cyberpsychology & Behavior*, 4:2 (2001), 307–315.

7   Short et al., *Social Psychology of Telecommunications.*

8   J Hauber, H Regenbrecht, M Billinghurst and A Cockburn, "Spatiality in Videoconferencing: Trade-offs between Efficiency and Social Presence," in *CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada, 4-8 November, 2006* (New York: ACM Press, 2006), 413–22.

9   S Benford, C Brown, G Reynard, and C Greenhalgh, "Shared Spaces: Transportation, Artificiality, and Spatiality," in *CSCW '96: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (New York: ACM Press, 1996), 77–86.

10   H Regenbrecht, M Haller, J Hauber and M Billinghurst, "Carpeno: Interfacing Remote Collaborative Virtual Environments with Table-Top Interaction," *Virtual Real.* 10:2 (2006), 95–107.

Jörg Hauber completed his PhD in Computer Science at the University of Canterbury, New Zealand in 2008. In his research he investigated human factors in advanced video-conferencing systems. Prior to his doctoral studies, Jörg obtained a Dipl-Ing(FH) Degree (Medical Engineering) from the University of Applied Sciences Ulm (Germany) as well as a MSc Degree (Information Technology and Automation Systems) at the Graduate School of the University of Applied Sciences Esslingen (Germany).