The Impact of Collaborative Style on the Perception of 2D and 3D Videoconferencing Interfaces

Joerg Hauber[1], Holger Regenbrecht[2,*], Andy Cockburn[3] and Mark Billinghurst[1]

[1]*HITLabNZ, Christchurch, New Zealand*

[2]*University of Otago, Dunedin, New Zealand*

[3]*University of Canterbury, Christchurch, New Zealand*

Abstract: Video-based collaborative virtual environments (CVE) attempt to emulate face-to-face meetings by immersing remote collaborators in a shared 3D virtual setting. To investigate potential advantages of this novel type of collaborative user interfaces for creating a better sense of social presence and affording a more efficient collaborative process we conducted an empirical study in which pairs of users solved a simple task (matching a set of celebrity photos with a set of quotes) using four different media: face-to-face, a standard desktop videoconferencing system (VC), a desktop video-CVE, and a stereo large-screen video-CVE. As expected, results showed that face-to-face provided a significantly stronger sense of social presence than any of the systems, but relatively little differences showed between the systems themselves. However, significant gender effects emerged in an ex-post analysis for the different system types, with females perceiving more social presence when using the standard video conferencing environment and less with the video-CVE conditions, while males showed the opposite effect. Linguistic analysis of audio transcriptions and video analysis further illuminates differences between collaboration styles of males and females across the collaborative conditions. We discuss the implications of our findings for future studies into CVEs and video conferencing systems.

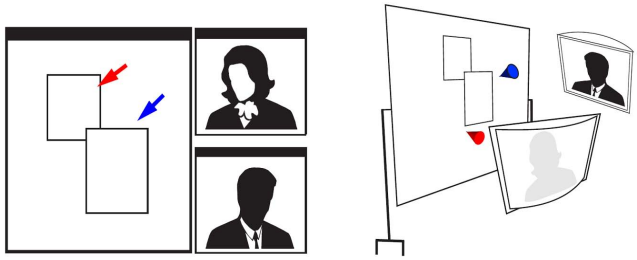
Keywords: Social presence, gender, spatiality, collaborative virtual environments, video-mediated communication, teleconferencing.

1. INTRODUCTION

The goal of 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. In this context, modern video-conferencing technology is commonly used to connect remote people who want to talk, work, or learn with each other. Video offers a visual communication channel that conveys several non-verbal communication cues such as facial expressions and gestures, which are absent in normal telephone calls. Still, compared to being face-to-face, video-communication feels cold, impersonal, unsociable, and insensitive [1], raising the question how it can be improved.

One shortcoming of video conferencing, compared to face-to-face collaboration, is the absence of a shared 3D frame of reference between the participants which reduces the sense of collocation, decreases gaze awareness, and thus may impede the establishment of a common collaborative context (see Fig. (**1**), standard VC).

Video-collaborative virtual environments (video-CVEs) are novel Video Mediated Communication (VMC) interfaces which seek to address these problems by re-introducing a virtual 3D context into which distant participants are mentally transported. Fig. (**1**) schematically depicts a video-CVE, seen from a 3rd person perspective. As can be seen in Fig. (**1**), users are represented by their own avatars through which they can then interact with the shared environment and with each other.



a) Standard VC b) Video-CVE

Fig. (1). Interface approaches to videoconferencing.

Although working prototypes of video-CVEs have demonstrated their technical feasibility, research into the value of video-CVEs for supporting remote collaboration is still in its infancy. In this article we present a study that explores how users collaborate within video-CVEs by directly comparing collaborative processes and user perceptions when mediated by four conditions: two video-CVEs systems, a standard video conferencing system, and a face-to-face condition. With this study we contribute to a better understanding of human factors in video-CVEs. Our work is significant, because it ties the area of collaborative virtual environments to the empirical body of classical cross-media comparisons found in HCI literature.

*Address correspondence to this author at the University of Otago, Information Science, P.O. Box 56, 9054 Dunedin / New Zealand;
Tel: ++64 (0)3 479 8322; Fax: ++64 (0)3 479 8311;
Email: holger.regenbrecht@otago.ac.nz

1874-107X/12

Table 1. Cross-Media Studies by Task and Conditions

	Conditions			
Task Type	**FtF**	**Video**	**Audio**	**References**
Trust and Deception		●	●	Wichman [10]
Negotiation	●	●	●	Short [12]
Source - Seeker task	●	●	●	Chapanis [3]
Brainstorming	●	●	●	Williams [49]
Discussion	●		●	Rutter and Stephenson [50]
Discussion		●	●	Rutter *et al.* [51]
Real project		●	●	Tang [18]
Discussion	●	●[a] ●[b]		Sellen [52]
Information Exchange	●	●[c] ●[d]		O'Conaill *et al.* [53]
Collaborative Design	●	●	●	Olson *et al.* [5]
Informer - Follower task		●[e] ●[f]	●	O'Malley *et al.* [6]
Informer - Follower task		●[g] ●[h]	●	Doherty-Sneddon *et al.* [7]
Decision making		●	●	Daly-Jones *et al.* [48]
Informer - Follower task		●	●	Veinott *et al.* [8]
Discussion of photos		●	●	de Greef and IJsselsteijn [20]
Informer - Follower task		●[i] ●[j]	●	Monk and Gale [15]
Trust and Deception	●	●	●	Bos *et al.* [11]

FtF = face-to-face. ● indicate conditions compared. [a]Hydra, spatially separate screen-camera unit for every participant; [b]Picture in Picture (PiP) video; [c]high quality video; [d]low quality video with delay; [e]video shows face only; [f] video shows head and shoulder; [g]video tunnel: eye contact possible; [h]video tunnel with offset: no eye contact possible; [i]video tunnel with full gaze awareness; [j]video tunnel with eye contact only.

We are addressing the following two main problems with the work presented here: (1) What are the potential benefits of spatial embodied interaction as used in video-CVEs in comparison to standard video conferencing systems? and, as an ex post problem statement: (2) Are there any gender differences in using standard and video-CVEs?

2. RELATED WORK

We first review several studies that compare collaboration across audio-only, audio-video, and face-to-face situations, describing both the measures applied and empirical findings. Then, we briefly review human factors research in the area of collaborative virtual environments, and finally provide a motivation for our study.

2.1. Cross-Media Comparisons

The basic rationale for adding video to audio seems straightforward: video adds some "value" compared to audio-only communication which improves the outcome, facilitates the process, and leads to greater satisfaction of telecommunication. In this sense, VMC resembles face-to-face more closely than audio-only communication.

A number of questions remain: How does the addition of video change a remote interaction? Does video always add value to audio? Is VMC always preferred over audio-only? Is the nature of VMC more similar to face-to-face or rather to telephone conversations? Compared with face-to-face conversations, what are the shortcomings of VMC? What differences emerge between different versions of VMC?

Various comparative user studies have addressed these and similar questions. In the remainder of this section the findings of these studies are summarised based on observed differences in product, process, and satisfaction measures (explained later).

Table **1** lists the studies reviewed. The media conditions focused on in this study are face-to-face (FtF), audio-video (AV), and audio-only (AO). Some of the studies comprised additional conditions which are neglected here to keep a uniform format. Some studies involved two different versions of the audio-video condition as indicated by two dots in the same field. Superscripts lead to descriptions of the different AV conditions in the bottom of the table.

In cross-media studies, participants collaborate on the same specially designed experimental task in different communication conditions. Experimenters then measure and investigate the differences in collaboration that surface in direct comparison. Gutwin and Greenberg [2] distinguish between product measures, process measures, and satisfaction measures:

- **Product measures** evaluate collaborative outcomes, considering both time and quality.
- **Process measures** examine the efficiency of collaborative activities by analysing speech and interaction patterns of participants.
- **Satisfaction measures** assess the quality of a communication medium based on the subjective opinion of the participants who used it during the experiment.

The following three sections combine and discuss the results of the studies listed in Fig. (**2**) with regard to each of these measures.

2.1.1.Product Measures

Product measures assume that differences in the communication media lead to measurable differences in the collaborative outcomes. The nature of these outcomes depends on the chosen experimental task. Tasks may have a "well-defined" goal that can be reached through collaboration (e.g. finding predefined locations in a street map [3]). For this type of task, the outcome can be measured in terms of the time it takes the team to reach the solution.

Other experiments applied "ill-defined" problems, that is, they do not have one fixed solution a priori [4]. Design tasks such as the one used in Olson *et al.* [5] fit into the category of ill-defined problems. In their study, participants were asked to design an automated post office. The outcome of the collaboration was then measured in terms of the quality of the final design, as assessed by a jury of experts.

Fig. (2). A set of quotes and photos. Participants had to collaborate to find matching pairs (solution: 0-A8, 1-A0, 2-A2, 3-A5, 4-A9, 5-A4, 6-A7, 7-A1, 8-A6, 9-A3).

Table 2. Cross-Media Studies: Product Measures and Results

Product Measures	Findings	References
Completion Time	FtF=AV=AO	Chapanis [3]
	AV=AO	Daly-Jones *et al.* [48]
Quality of Outcome	FtF=AV=AO	Williams [49]
	FtF=AV>AO	Olson *et al.*[5]
Reproduction Accuracy	$AV^{e}=AV^{f}=AO$	O'Malley *et al.* [6]
	$AV^{g}=AV^{h}=AO$	Doherty-Sneddon *et al.* [7]
	AV>AO *	Veinott *et al.* [8]
	AV=AO **	Veinott *et al.* [8]
	$AV^{i}=AV^{j}=AO$	Monk and Gale [15]
	AV≠AO	Wichman [10]
Negotiation Results	FtF=AV≠AO	Short [12]
	FtF=AV=AO	Bos *et al.* [11]

FtF=face to face; AV=audio-video; AO=audio-only. The symbol ≠ indicates a significant difference found; The symbol > indicates significant difference or trend found; = indicates no significant difference found. [e]video shows face only; [f] video shows head and shoulder. [g]video tunnel: eye contact possible; [h]video tunnel with offset: no eye contact possible; [i]video tunnel with full gaze awareness; [j]video tunnel with eye contact only; *participants were non-native English speakers; ** participants were native English speakers.

Another frequently used experimental task is the "Map Task" [6-8]. Here, two participants get two slightly different versions of a terrain map. The map of one participant (informer) shows an additional path that has to be reproduced as accurately as possible by the other participant (follower). Reproducing the path accurately is challenging and can only be achieved by means of effective communication. The deviation between the actual and the ideal path therefore serves as a product measure for the effectiveness of communication and thus of the quality of collaboration.

Experimenters also applied social dilemma games such as the "Prisoner's Dilemma (PD)" [9] to investigate the impact of a communication medium on a participant's level of trust and willingness to cooperate. In the PD game, every participant has the repeated choice to cooperate with or betray the other participant. If both players decide to cooperate, both receive a pay-off as a reward. However, if player A successfully betrays player B (player B chooses to cooperate while player A defects), he receives a higher pay-off. If both players defect, neither receives a pay-off. Every player's ultimate goal is to maximise his/her pay-offs over the course of several rounds. Table **2** shows the product measures applied and the results obtained in the studies considered in this review of cross-media studies.

As can be seen in Table **2**, the time needed to complete a well-defined task is not affected by the different communication media. The same holds for the quality of the obtained results in ill-defined tasks, with the exception of the results reported by [5], who found a marginally worse quality in the design solutions that were created in the audio-only conditions.

Also the repeatedly applied Map Task did not produce reliable results, since the accuracy of the reproduced path did not differ across various audio-video and audio-only conditions. One noticeable exception was a study reported by [8] who found significantly better results in the audio-video condition if participants were non-native English speakers.

In contrast, studies involving a negotiation task produced more reliable differences. Wichman [10] applied the Prisoner's Dilemma game and found a significantly higher percentage of cooperation in the audio-video condition compared to audio only.

This result could not be reproduced by Bos *et al.* [11], who compared the outcomes of thirty rounds of a PD-like task between face-to-face, audio-video, as well as audio-only conditions. However, their results showed that the level of cooperation was initially lower in the mediated conditions and only slowly converged with the face-to-face level over time. This suggests that establishing trust takes longer if communication is mediated. The authors explicitly mention that to their own surprise the outcomes of audio-only and audio-video conditions were almost identical over the whole course of the thirty rounds.

Short [12] asked participants to negotiate a fictitious situation, where one participant represented a point-of-view he or she was allowed to choose before the experiment, while another participant then had to take the opposing viewpoint (independent of whether that view was consistent with his or her true beliefs).

Short found that compared to the audio-only condition, the results of the negotiations were more in favour of the consistent views, that is, where the represented view matched the true belief of a participant, when participants negotiated in the face-to-face or the audio-video condition.

As can be seen from these results, apart from tasks that involve negotiation or trust, product measures rarely discriminated between face-to-face, video-mediated, and audio-only collaboration, even if tasks were specially designed to bring forth the assumed benefit of video over audio. Monk *et al.* [13] argue that in an experimental situation, people tend to protect their primary task (getting the work done) at a cost to any secondary task or to subjective effort. For cross-media comparisons this implies that participants always focus on

Table 3. Cross-Media Studies: Process Measures and Results

Process Measures	Findings	References
Number of Words	$AV^{e}=AV^{f}>AO$	O'Malley *et al.* [6]
	$AV^{g}>AV^{h}=AO$	Doherty-Sneddon *et al.* [7]
	FtF=AO	Rutter and Stephenson [50]
	$AV^{i}<AV^{j}=AO$	Monk and Gale [15]
Number of Turns	$FtF=AV^{c}>AV^{d}$	O'Conaill *et al.* [53]
	$FtF=AV^{a}=AV^{b}$	Sellen [52]
	AV=AO	Daly-Jones *et al.* [48]
	$AV^{e}=AV^{f}>AO$	O'Malley *et al.* [6]
	$AV^{g}>AV^{h}=AO$	Doherty-Sneddon *et al.* [7]
	AV=AO	Veinott *et al.* [8]
	FtF=AO	Rutter and Stephenson [50]
Turn Length	$FtF=AV^{c}<AV^{d}$	O'Conaill *et al.* [53]
	$FtF=AV^{a}=AV^{b}$	Sellen [52]
	AV=AO	Daly-Jones *et al.* [48]
	FtF=AO	Rutter and Stephenson [50]
Switching Times	$FtF<AV^{a}=AV^{b}$	Sellen [52]
Handover by Name	$FtF=AV^{c}<AV^{d}$	O'Conaill *et al.* [53]
Overlapping Speech	FtF=AV>AO	Rutter *et al.* [51]
	$FtF=AV^{c}=AV^{d}$	O'Conaill *et al.* [53]
	$FtF>AV^{a}=AV^{b}$	Sellen [52]
	AV=AO	Daly-Jones *et al.* [48]
	FtF>AO	Rutter and Stephenson [50]
Interruptions	FtF=AV>AO	Rutter *et al.* [51]
	$FtF=AV^{c}=AV^{d}$	O'Conaill *et al.* [53]
	$FtF>AV^{a}=AV^{b}$	Sellen [52]
	$AV^{e}=AV^{f}>AO$	O'Malley *et al.* [6]
	$AV^{g}=AV^{h}>AO$	Doherty-Sneddon *et al.* [7]
	FtF>AO	Rutter and Stephenson [50]
Relative time spent . . .		Olson *et al.* [5]
. . . clarifying issues	FtF=AV<AO	
. . . clarifying what was meant	FtF<AV=AO	
. . . managing the meeting	FtF<AV<AO	

FtF=face-to-face; AV=audio-video; AO=audio-only. The symbols > and < indicate significant differences or trends found; = indicates no significant difference found. [a]Hydra, spatially separate screen-camera unit for every participant; [b]Picture in Picture (PiP) video; [c]high quality video; [d]low quality video with delay; [e]video shows face only; [f] video shows head and shoulder; [g]video tunnel: eye contact possible; [h]video tunnel with offset: no eye contact possible; [i]video tunnel with full gaze awareness; [j]video tunnel with eye contact only;

delivering the best possible product, even if different media conditions demand more or less effort to achieve it. Product measures may therefore not be the best means to assess the value of a telecommunication medium for collaboration.

2.1.2. Process Measures

Process measures investigate the differences in speech and interaction patterns that emerge during the course of collaboration. The underlying assumption is that different communication conditions afford different grounding mechanisms which in turn lead to differences in verbal conversation styles and content. Process measures can be obtained in real time by observation, or, more typically, by in-depth analysis of audio and video recordings and extracted transcripts. Examples for typical process measures are the number of spoken words, speaker turn taking behaviour such as turn frequency and length, overlapping speech, interruptions, or the number of questions.

Table **3** lists several process measures along with the results found in the studies reviewed.

The process measures reported in the reviewed studies did not produce consistent results. Apparently, the influence of the nature of the experimental task caused a greater variance in results than that caused by the different communication conditions, making it hard to compare measures of the same kind between two or more studies. However, there seems to be consent on some general characteristic tendencies that were repeatedly observed.

Face-to-face communication is spontaneous with frequent speaker changes, frequent interruptions, and overlapping speech.

In contrast, video-mediated and audio-only conversations are more formal and rigid, characterised by fewer, but longer "lecture-like" turns, hindered turn switching, fewer interruptions, and less overlapping speech. O'Conaill *et al.* [14] found the formal character to be particularly apparent if the mediated audio suffers from a delay.

Olson *et al.* [5] analysed the content of spoken turns and found that people in face-to-face situations devoted fewer of their spoken turns to clarification and coordination purposes than people whose conversations were mediated. Between the two mediated conditions, people communication via an audio-video link also used fewer turns for clarification and managing the meeting than people using audio-only communication. The smaller verbal overhead in the face-to-face and audio-video conditions suggest that participants used visual cues during their conversations which allowed their verbal conversations to be more task-focused and thus more efficient.

Monk and Gale [15] also demonstrated that the provision of full gaze awareness in video communication could reduce the number of words spoken to one half. They see this reduction as a clear sign of increased communication efficiency and superiority of that type of video communication. Other studies, however, yielded contradicting results. Doherty-Sneddon *et al.* [7] and O'Malley *et al.* [6] for example report face-to-face and video conditions to be wordier than audio-only. Interpreting the number of spoken words solely in terms of efficiency is therefore not conclusive. A possible explanation for these inconsistent findings is that the maps of the "map task" used by Doherty-Sneddon *et al.* and O'Malley *et al.* offered clear verbal referents making it easy to verbalise objects (e.g. monuments, lakes, forests, etc.). Monk and Gale on the contrary used more abstract pictures that were much more challenging to explain and refer to verbally (e.g. an electron microscope slide showing more than 100 identical benzene molecules).

If the cost of verbal grounding is high initially, as in the latter case, there is a higher motivation for participants to use cheaper alternatives. Then, the provision of additional visual cues can substantially change the process of collaboration, because participants frequently shift from the verbal to the visual channel to reduce their collaborative effort. In contrast, if the initial cost of verbal grounding is low, like in the map task, providing additional visual cues, does not necessarily lead to a more effective way for grounding, but gives the communication a more social and personal character. And, as Doherty-Sneddon *et al.* mentioned, people therefore talk more when they feel more satisfied and comfortable in a certain communicative situation. Examples like these illustrate the limited external validity of the results obtained and underline the importance of taking the circumstances of the individual task into consideration.

Table 4. Cross-Media Studies: Satisfaction Measures and Results

Satisfaction Measures	Findings	References
Preference	AV>AO	Daly-Jones *et al.* [48]
	AV>AO	Tang [18]
	AV[cd]>AO	O'Conaill *et al.* [53]
	FtF>AV>AO	Olson *et al.* [5]
Subj. Comm. Efficiency	AV>AO *	Veinott *et al.* [8]
	AV<AO **	Veinott *et al.* [8]
Subjective Quality of Outcome	FtF>AV>AO	Olson *et al.* [8]
Subjective Effort	AO>AV	Daly-Jones *et al.* [48]
Interpersonal Awareness	AV>AO	Daly-Jones *et al.* [48]
Ability to take control of the conv.	FtF>AV[a]=AV[b]	Sellen [52]
Subjective Interactiviy	FtF>AV[a]=AV[b]	Sellen [52]
Selective Attention	FtF>AV[a]>AV[b]	Sellen [52]
Knowing when others were listening	FtF>AV[a]=AV[b]	Sellen [52]
Social Presence	AV>AO	de Greef and IJsselsteijn [20]
	FtF>AV[e]>AV[f]	Hauber *et al.* [21]

FtF=face-to-face; AV=audio-video; AO=audio-only. The symbols > and < indicate significant differences or trends found; = indicates no significant difference found. [a]Hydra, spatially separate screen-camera unit for every participant; [b]Picture in Picture (PiP) video; [c]high quality video; [d]low quality video with delay; [e]shared spatial context; [f]WYSIWIS interface; * Participants were non-native English speakers; **Participants were native English speakers.

Consistent within all results of process measures is the fact that whenever a difference between face-to-face and audio-only emerged, the score of the audio-video condition was found somewhere in-between. This places video-mediated communication between audio-only and face-to-face communication. Yet, Sellen [16] and Williams [17] see a bigger resemblance between video-mediated conversations and audio-only communication rather than between video-mediated communication and face-to-face conversations.

2.1.3. Satisfaction Measures

Satisfaction measures assess the quality of communication based on the user's subjective experience. Participants who were exposed to a communicative situation are typically asked to answer a set of questions which tap into several dimensions of interest. These questions can either be presented in the form of questionnaires or can be asked orally in interviews. Satisfaction dimensions include the perceived performance, perceived effort, comfort level, perceived social presence, perceived workspace awareness, or enjoyment. In experiments where each participant gets exposed to more than one communication condition, the focus is on the perceived differences between them. The experimenter may therefore ask participants to rank the involved conditions according to their preferences. A selection of satisfaction measures along with their results is shown in Table **4**.

The results obtained from satisfaction measures seem to be consistent: people clearly favour face-to-face over audio-video, and audio-video over audio-only communication. This applies for measures of preference, effort, awareness, control, and social presence. The only noticeable exception is reported by Veinott *et al.* [8], who found that a group of native English speakers perceived audio-only to be more efficient than audio-video, very much to the experimenters' own surprise.

Tang [18] (1992) reported clear evidence for the benefit of adding video to audio. He conducted a field study, observing a real project team of four (later five) members over the duration of fourteen weeks. During that time, two new teleconferencing systems were introduced to the team, one of which offered the possibility for real-time video. Observation of actual system usage in comparison with other standard media like email or phone, complemented by interviews with the team members, revealed that availability of video was the key factor for system usage and system preference. When interviewed, the team members pointed out several benefits the video realised. Video facilitated the communication as gestures could be used. Also, while talking, users could see each other´s reactions and instantly monitor if they were being understood. Longer speech pauses, which are hard to interpret in audio-only were "demystified" by the video, because remote participants were aware of activities in the background that prohibited the other partner from talking. The members even noticed that being able to see the others lead to an increased engagement in social, personal contact through video, which ultimately improved the communication and awareness among the team.

The participants in the study conducted by O'Conaill *et al.* [14] stated similar advantages of video compared to audio-only. Being able to know who was at the remote location was seen as a clear benefit which would also foster the feeling of "not talking into the void".

Participants of other studies also rated video to lead to more efficient communication involving less effort while offering a higher level of control, awareness, and social presence. Participants in the study conducted by Olson *et al.* [5] furthermore felt their collaborative outcome to be superior compared to the resulting outcome of the audio-only condition. This is particularly interesting as the participant´s subjective opinion was not confirmed by the expert jury.

Social presence theory [19] proposes a more elaborate way of further quantifying participants' subjective attitudes towards a communications medium. Social presence conveyed by a medium is assessed through a set of semantic differentials including insensitive--sensitive, impersonal--personal, cold--warm, and unsociable--sociable. Media that support more non-verbal communication channels are higher in social presence because they are typically perceived as warmer, more personal, more sociable, and more sensitive. Assessed in that way, social presence increased for example, when video was added to a groupware application that allowed remote participants to view and discuss photos [20]. Hauber *et al.* [21] demonstrated that social presence of VMC can be further increased if a sense of spatiality is maintained in a remote encounter.

In all the studies that included a face-to-face condition, participants always clearly preferred that over any form of mediated communication.

2.1.4. Summary of Cross-Media Studies

Evaluating video-mediated communication is no trivial undertaking. Many factors and subtleties have to be taken into consideration which may distort the results of a study. This makes it hard to compare the results of different studies that included the same communication conditions, but used different tasks and participants.

The quality of a communication medium cannot be observed directly, but has to be derived from a set of measures which examine the outcome of communication, the process of communication, or subjective user satisfaction. Product measures are sensitive only to gross changes and therefore frequently fail to picture any media differences. Process measures are very time-consuming to collect, but are able to identify differences between the interaction patterns that different media bring forward. They are sensitive to the experimental task and the type of documents that are involved, and should therefore always be interpreted and compared with caution. Finally, satisfaction measures produce the most reliable results, both in sensitivity and cross-study concordance.

Based on all the collected results in the reviewed studies the following three main points can be concluded:

- Video can add value to audio: the degree to which video is beneficial in terms of better outcomes or communication efficiency is first and foremost determined by the type of the collaborative task.
- Good audio is more important than good video: most of the studies were conducted in a controlled environment with ideal conditions which allowed for quality audio and quality video transmission. However, especially the study conducted by O'Conaill *et al.* [14] made clear that any fluidity and efficiency of communication processes breaks down immediately with poor and delayed audio. Any expected advantages through the addition of video rely on accurate timing and synchronicity between video and speech and therefore presupposes the maintenance of high quality audio with minimal delay. The quality of audio should therefore never be compromised for higher video quality [22].
- People like video: the satisfaction measures revealed that the people in the studies all liked to have video, mainly because it provided basic awareness information and allowed them to monitor facial expressions and other non-verbal reactions in the course of a remote conversation.

Though, there is only a limited body of work on gender aspects in videoconferencing, for instance Wheeler [23] reported that women are less self-conscious when using videoconferencing technology but have a favourable attitude towards it. In contrast, Maurin *et al.* [24] found that male paramedics have more favourable attitudes to collaborate with a remote physician than females. Stuhlmacher *et al.* [25] found that women are significantly more aggressive

when using computer-mediated communication technologies in comparison to face-to-face communication. This is in line with research by Wachter [26] who found that women feel more able to dominate the partner in a videoconferencing condition. Finally, deGreef *et al.* [20] found a gender interaction in which a talking head video conveyed more social presence for female participants than for male participants.

Recent work by Teoh *et al.* [27, 28] found significant gender differences in videoconferencing in relation to trust and the availability of body language. Other recent studies, like Nguyen and Canny [29] present the gender distribution in their sample but do not report on specific differences found. Lowden and Hostetter [30] reported that female participants who used videoconferencing in high frequency were significantly more satisfied than males. Also, gender had a positive impact on social presence.

However, to our knowledge there is no comprehensive study on gender in collaborative virtual or video-mediated environments.

3. COLLABORATIVE VIRTUAL ENVIRONMENTS CVE

With the emergence of the internet and the feasibility to link distant computers, the question arose whether VR technology could, if used in a multi-user setting, not only create a sense of "being there" in a different space, but also induce a sense of "being together" with others in that space. The idea of Collaborative Virtual Environments (CVEs), that is, virtual worlds shared by participants across a computer network, was born.

The technology of CVEs aims to transform today's computer networks into navigable and populated 3D spaces that support collaborative work and social play [31]. The emergence of CVEs can be seen as the result of a convergence of research interests within the VR and CSCW communities.

All CVEs share the key characteristics of spatial immersion and embodiment that set them apart from many other collaborative systems. Spatial immersion refers to the fact that CVEs present an egocentric perspective of the virtual scene to the user, suggesting that they are an active element of the virtual scene rather than a person on the outside looking in. The actual view into a virtual scene is controlled by the user, where a shift of the geometrical origin of the view is perceived as self movement and any rotation about the view axis can be understood as a change in one's gaze direction. Coupled with the subjective view into a scene and the change thereof is the experience of being present at a fixed location and orientation within the virtual environment (VE).

To make that subjective position and orientation perceivable for others who are immersed in the same VE, users are represented through virtual embodiments or "avatars", which appear at the very virtual location from which the user experiences the VE. Avatars vary in appearance, ranging from simple geometrical shapes to fully animated realistic humanoid representations. Users "see" and "hear" through the eyes and ears of their avatars. They therefore treat the avatars as if they were the user they are representing. This allows for a quasi-direct social interaction situated in the VE.

Table 5. User Studies in CVE Research

Focus of Study	References
Avatar appearance	Parise *et al.* [28]
	Nowak and Biocca [29]
	Garau *et al.* [30]
	Bailenson *et al.* [31]
Usability inspection	Greenhalgh and Benford [32]
	Normand *et al.* [33]
Turn-taking behaviour	Bowers *et al.* [34]
Small group dynamics	Slater *et al.* [35]
Cross-media comparisons	Nakanishi *et al.* [36]
	Sallnäs [37]

3.1. Background of Study

Some user studies have been conducted in an attempt to investigate human factors when people interact with others in CVEs. In the absence of a dedicated CVE evaluation methodology, the scope and methods applied in these studies vary considerably borrowing from single-user VR evaluations, general usability assessments to communication analysis. Table **5** lists a selection of studies that will be briefly reviewed in the following paragraphs.

The impact of different avatar appearances on social interaction is among the most studied design parameters in CVE research. Parise *et al.* [32] for example investigated how the level of cooperation in a social dilemma game was influenced by the more or less realistic and human-like avatar representation of the participants. They found higher levels of cooperation and trust when participants were represented by human-like avatars compared to a control condition that embodied interlocutors as talking dogs. However, Nowak and Biocca [33] found people preferred a less anthropomorphic representation of others over highly anthropomorphic avatars. They observed that an avatar which appears too realistic may easily lead to disappointment and mistrust if high expectations with regards to the avatar's behaviour that are fostered by its realistic appearance are not met. Garau *et al.* [34] and Bailenson *et al.* [35] also explored the relation of pictorial realism (how real avatars look) and behavioural realism (how human-like avatars behave). Garau *et al.* found that realistically looking avatars with higher behavioural realism (controlled by controlled gaze) outperformed realistically looking avatars with low behavioural realism (controlled by random gaze). In summary, these results suggest that both pictorial and behavioural realism have to be carefully balanced to design avatars for predictable, efficient and enjoyable avatar-mediated communication.

Greenhalgh and Benford [36] and Normand *et al.* [37] report first experiences with the research prototypes MASSIVE and DIVE based on informal user observations. Common usability issues that came to attention were user problems with navigation and issues due to the limited field of view. In the absence of peripheral vision, a group of users in MASSIVE had, for example, problems forming a circle with their avatars.

Bowers *et al.* [38] focused on turn taking behaviour in a CVE and concluded that users utilise their embodiments in systematic ways to resolve or anticipate turn-taking problems. Users frequently moved their avatar around and positioned them in order to become "face engaged" with other avatars they wished to interact with.

Slater *et al.* [39] compared group dynamics between a face-to-face and a CVE condition. Teams of three persons worked on a collaborative paper-puzzle task. In the CVE condition, one of the three was equipped with a head-mounted display (HMD), while the others saw their view into the virtual scene through a less immersive desktop computer monitor. The person wearing the HMD emerged as the group leader significantly more often than the others, indicating that immersion enhances leadership capability.

Only few CVE-studies involved comparisons of VMC and avatar-mediated communication. Nakanishi [40] compared the speech and motion patterns of groups of seven members who worked on three different conversational tasks in a standard video-conferencing condition, to their FreeWalk video-CVE, and to a non-mediated face-to-face scenario. Results revealed that there were significantly more conversational turns in FreeWalk compared to both standard video and face-to-face conditions and that participants moved around more in FreeWalk than when being physically co-present.

A second example is a series of two studies reported by Sallnäs [41] who investigated the effects of communication mode on social presence, virtual presence, and performance in CVEs. In the first experiment, teams of two participants met in the online 3D-world "ActiveWorlds" and communicated through either text-chat, audio, or an audio-video link which was established on a second PC. Social presence, virtual presence, and the number of exchanged messages were lower in the text-chat condition than in the audio- or video-conferencing condition. Participants also spoke fewer words in the video-conference condition. In the second experiment, Sallnäs compared collaboration in a CVE audio- and CVE video-condition with collaboration in a Web audio-conference and a Web video-conference condition. Participants spent more time in the video than in the audio conditions, and spoke more words per second in the Web conditions. Sallnäs concluded that both the communication media used and the collaborative environment have an impact on user experience and users' communication behaviour.

These examples demonstrate some of the dimensions that have been of interest to researchers in the field so far. However, CVE research is still in its infancy if one considers the rather exploratory nature of most of the studies as well as the heterogeneity of the measures that were used.

3.2. Motivation for this Study

CVEs are fast advancing platforms for remote collaboration. However, the question of the potential benefit of spatial embodied interaction compared to state-of-the-art VC remains unclear. With this study we are addressing this gap. By following the comparative research approach of cross-media studies we believe that valuable lessons can be learnt on how video-CVEs influence collaboration.

The advantage of comparing video-CVEs with a standard VC tool is that users are embodied in a very similar way -- namely by their video representations. It thus allows us to ascribe observed differences to our dimensions of interest without having to worry about interference caused by otherwise crude and artificial looking avatars common to some other CVEs.

4. EXPERIMENT DESIGN

We designed a collaborative pair-matching task for this experiment. Ten photos of well-known personalities as well as ten significant quotes were presented to participants (see Fig. (**2**) for an example). The task for the team of two participants was to find as many correct celebrity-quote pairs as possible.

Five sets, each consisting of ten celebrities and quotes, were compiled for the experiment. Every set contained a broad mix of celebrities, including philosophers, musicians, or fictional characters. We trialled the sets during an open house demo with 60 participating visitors. The so gathered results allowed us to get an estimate of which of the quotes were considered easy and which were not. Where necessary, we swapped quotes between sets accordingly to equalise the difficulty among all sets.

Every photo had its own ID, every quote had its own tag, as displayed in the little boxes. Participants could enter an ID and tag of a celebrity--quote pair into our special mapping application, the "Map-o-Mat" (see Fig. (**3**), left). To finish

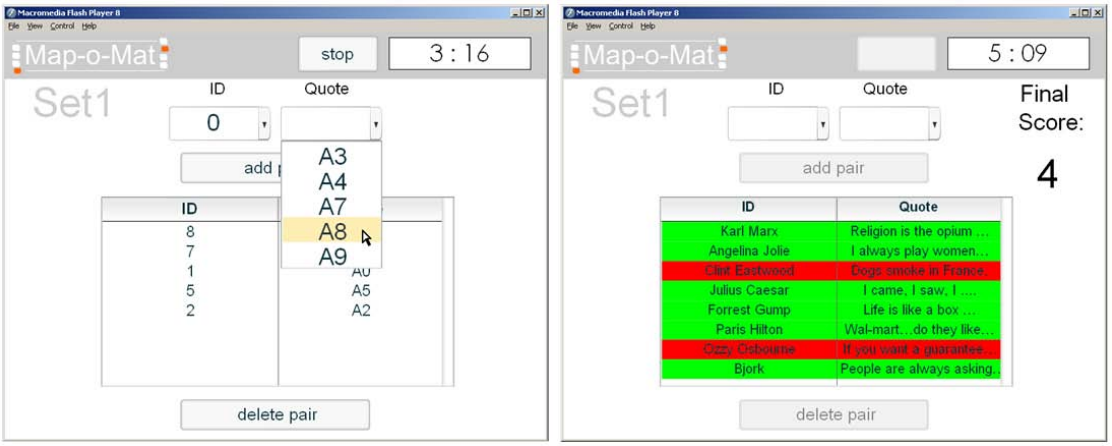


Fig. (3). The "Map-o-Mat" mapping application. It allowed participants to enter celebrity--quote pairs.

Fig. (4). Experiment Conditions.

the task, they could press a stop button which immediately processed the entered items and displayed the results of that round (see Figure 3, right). The final score was calculated as the number of correct pairs minus the number of incorrect ones.

As we will later see significant gender differences surfaced in an ex-post analysis. Unfortunately, we did not foresee this strong effect and hence did not design for it. We (intuitively) recruited same gender pairs of participants though.

4.1. Experiment Conditions

The study was designed as a within-subjects experiment involving the four media conditions depicted in Fig. (**4**).

4.1.1. Condition "Face-to-Face" (FtF), (See Fig. (4) Top Left):

Participants collaborated in the same room. Two A4 sheets of paper showing photos and quotes were attached on the opposite ends of a long table (2.4 m), oriented sideways, so that they could not be looked at by one person at the same time. The Map-o-Mat application ran on a laptop computer located in the middle of the table (not visible in the photo).

4.1.2. Condition „Standard Video-Conferencing" (sVC), (See Fig. (4), Top Right):

Participants were located in separate rooms and were connected with the commercially available video-conferencing software "Marratech 6.1", provided by Marratech for this experiment. The interface showed the video streams of both participants and a shared presentation area where celebrities and quotes were displayed on different slides. Each participant individually controlled which slide he or she wanted to look at. Telepointers could be used to point out certain details on the slides to the other participant, provided that the participant was looking at the same slide at that moment. The Map-o-Mat ran in a separate, shared application window (UltraVNC, version 10.1.1.3), allowing both participants to enter pairs.

4.1.3. Condition "Video-CVE Desktop" (vCVE_Desk), (See Fig. (4), Bottom Left):

Participants were located in separate rooms and met virtually in a video-CVE [42] that contained photos, quotes and the Map-o-Mat terminal. In this condition, the virtual room was presented to the participants on a standard 19" flat screen monitor placed on the desktop in front of them. Participants could control both the Map-o-Mat application and telepointers on the billboards using a standard computer mouse. In addition, a commercially available spatial controller device (SpaceMouse) was provided as a state-of-the-art means for navigation. Participants typically operated the space mouse with their left hand concurrently with the normal computer mouse in the right hand.

4.1.4. Condition "Video-CVE Immersive" (vCVE_im)

This condition was similar to condition "vCVE_desk", with the difference that the virtual environment was displayed as an 80" stereo-projection. The stereo image was created with a DepthQ "In Focus" stereo projector in combination with shutter glasses worn by the participants. To prevent discomfort for inexperienced users, we chose a low virtual eye separation in the stereo settings.

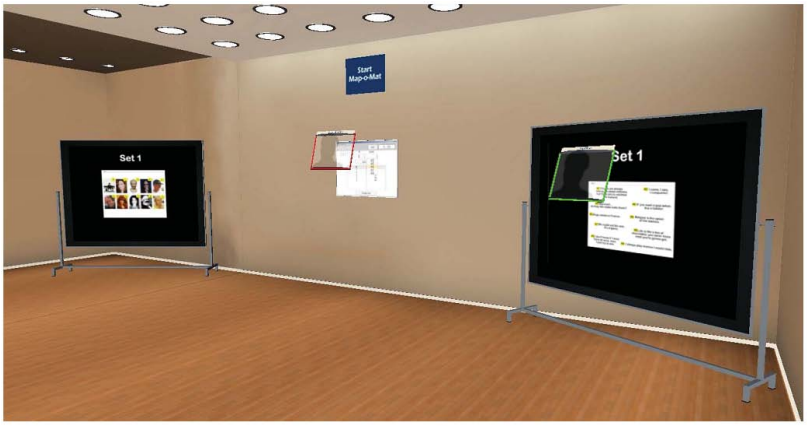


Fig. (5). The virtual room used in the experiment. Two billboards show the quotes and photos; an integrated terminal shows the Map-o-Mat application. As can be seen by the position of the avatars in the figure, one participant is entering some data into the Map-o-Mat (centre), while the other participant is reading some quotes on the billboard (right). Note: the avatars appear transparent because they are facing away from the observer.

Fig. (**5**) shows the virtual room used in the two video-CVE conditions. The three artefacts ("photos", "quotes", and "Map-o-Mat") were spatially separated so that participants had to actively navigate between them to get a closer look. The experience of "self-motion" in a virtual environment is a contributing factor to a sense of presence. Also, the changing position and orientation of an avatar provides a visual awareness mechanism for the other participant, who can automatically infer from the current location of the other avatar what the other person is looking at at that time.

In order to equalise the introduced effort for navigating between the quotes and photos in the video-CVEs, twelve additional "buffer slides" were also added between a photo-slide and a quote-slide in the standard video-conferencing condition. Participants could flick through the slides using the *PageUp* and *PageDown* button in about the same time it took them in the video-CVE conditions to get from one billboard to the other (about three seconds).

The PCs used to run the video-conferencing software at both ends were two identical, standard computers equipped with high end graphics cards. All computers involved in the experiment setup were connected via a 1 Gb Ethernet switch. Both work stations were equipped with standard USB webcams (CIF resolution) and standard teleconferencing headsets. To exclude effects based on differences in audio quality between conditions, the Marratech audio connection was also used in the two video-CVE conditions.

Two DV cameras with external microphones captured the video and audio of participants during the experiment.

4.2. Participants

The participants were post-graduate students from various departments at the University of Canterbury.

We opted for same gender teams of subjects who already knew each other. Therefore, we asked every participant to bring along a same-gender friend as his or her team partner. Participants were also required to be fluent in English.

Thirty six volunteers (26 male and 10 female) participated in the experiment. In 18 sessions, teams of two took part in four trials, for a total of 144 trials. The age of the participants ranged from 22 to 36 years (median age 26 years). Participants had no prior knowledge of the experiment except for the fact that the objective was to compare video-conferencing systems.

4.3. Procedure

For every one-hour session, two subjects were present. Upon arrival, the participants were asked to read and sign the Participant Information sheet, which outlined (1) the goal of the experiment, (2) the general procedure and (3) the anonymity policy of the experiment. Additionally, a short questionnaire collected demographic data.

After the task was explained to the participants, the first of four rounds (FtF, sVC, vCVE_desk, vCVE_im) began. The order of the four conditions and the five quote-celebrity sets used were controlled beforehand following a Latin square scheme. Before each round, participants had the opportunity to get familiar with the interface and to practice with the controls that were explained and demonstrated to them. To make sure that operating a SpaceMouse navigation device does not lead to problems during the actual experiment, participants could learn to master the SpaceMouse's functionality using a tutorial application provided by the manufacturer of the device. No advice on the general strategy of how to find matching pairs was given. Once the participants signalled that they had understood the interface and felt confident operating it, the actual sets of photos and quotes were loaded and the Map-o-Mat was started. This officially started the round. It was now up to the participants to identify the celebrities on the photo board and discuss and enter possible matching pairs into the Map-o-Mat application. Once a team decided not to enter any more pairs, one of the members had to hit the stop button on the Map-o-Mat application which marked the end of the round.

After each round, subjects were brought back into the same room and were asked to fill out the experiment questionnaire. After the fourth and final round they were briefly interviewed about their experience with the different interfaces and were asked to give their personal ranking of all four conditions. The total time for one session was approximately one hour.

Table 6. LIWC2001 Test Dimensions and Examples

Dimension	Abbrev	Examples	# Words
I. STANDARD LINGUISTIC DIMENSIONS Word Count	WC		
Words per sentence	WPS		
Sentences ending with ?	Qmarks		
Unique words (type/token ratio)	Unique		
% words captured, dictionary words	Dic		
% words longer than 6 letters	Sixltr		
Total pronouns	Pronoun	I, our, they, you're	70
1st person singular	I	I, my, me	9
1st person plural	We	we, our, us	11
Total first person	Self	I, we, me	20
Total second person	You	you, you'll	14
Total third person	Other	she, their, them	22
Negations	Negate	no, never, not	31
Assents	Assent	yes, OK, mmhmm	18
Articles	Article	a, an, the	3
Prepositions	Preps	on, to, from	43
Numbers	Number	one, thirty, million	29
II. PSYCHOLOGICAL PROCESSES Affective or Emotional Processes	Affect	happy, ugly, bitter	615
Positive Emotions	Posemo	happy, pretty, good	261
Positive feelings	Posfeel	happy, joy, love	43
Optimism and energy	Optim	certainty, pride, win	69
Negative Emotions	Negemo	hate, worthless, enemy	345
Anxiety or fear	Anx	nervous, afraid, tense	62
Anger	Anger	hate, kill, pissed	121
Sadness or depression	Sad	grief, cry, sad	72
Cognitive Processes	Cogmech	cause, know, ought	312
Causation	Cause	because, effect, hence	49
Insight	Insight	think, know, consider	116
Discrepancy	Discrep	should, would, could	32
Inhibition	Inhib	block, constrain	64
Tentative	Tentat	maybe, perhaps, guess	79
Certainty	Certain	always, never	30
Sensory and Perceptual Processes	Senses	see, touch, listen	111
Seeing	See	view, saw, look	31
Hearing	Hear	heard, listen, sound	36
Feeling	Feel	touch, hold, felt	30
Social Processes	Social	talk, us, friend	314
Communication	Comm	talk, share, converse	124
Other references to people	Othref	1st pl, 2nd, 3rd per prns	54
Friends	Friends	pal, buddy, coworker	28
Family	Family	mom, brother, cousin	43

Table 6. cont…

Dimension	Abbrev	Examples	# Words
Humans	Humans	boy, woman, group	43
III. RELATIVITY Time	Time	hour, day, oclock	113
Past tense verb	Past	walked, were, had	144
Present tense verb	Present	walk, is, be	256
Future tense verb	Future	will, might, shall	14
Space	Space	around, over, up	71
Up	Up	up, above, over	12
Down	Down	down, below, under	7
Inclusive	Incl	with, and, include	16
Exclusive	Excl	but, except, without	19
Motion	Motion	walk, move, go	73

4.4. Measures

Social presence was assessed for all four media conditions following the semantic differential technique proposed by Short [19]. We used the following seven bipolar pairs: cold -- warm, impersonal -- personal, insensitive -- sensitive, unsociable -- sociable, unpleasant -- pleasant, spontaneous -- formal and negative -- positive. In addition, the sense of physical presence was measured in the two video-CVEs using six items from the Igroup Presence Questionnaire (IPQ) [43]. Preference rankings were collected during the interviews with the participants.

Besides these subjective measures, audio transcripts were created for a subset of participants to investigate speech patterns and to provide grounds for interpretation of the subjective data. We further complemented the transcription analysis with a video analysis, which explicitly investigated the view coordination in the four conditions.

4.4.1. Linguistic Features Analysis

Extracted transcripts were analysed with the software "Linguistic Inquiry and Word Count LIWC2001[1] " (also see [44]). This program categorises words given in any text file with respect to more than 70 linguistic dimensions. The categorisation is based on a built-in dictionary that was developed and validated by linguists. Table **6** gives an overview of the test dimensions included in the dictionary. The output of a LIWC2001-analysis of a given text provides an overview of standard linguistic dimensions such as the total number of words, the number of words per sentence, as well as the relative usage of words belonging to the different word categories.

In addition to the default dictionary, a second dictionary was created where words and categories that were of particular interest or unique to this experiment were defined. This dictionary included categories for local and remote deixis, laughter, and the different types of quote references.

The relative usage of words belonging to the same word category were averaged for every condition, and then tested for main effects and interactions across the four conditions in a Mixed Model ANOVA.

Since the transcription of conversations is very time consuming, transcripts of only a subset of all teams were created and analysed following this methodology. The plan before the experiment was to transcribe and analyse the conversations of the last five teams. However, after the experiment, the subset was extended to the last eight teams in order to gather data from a balanced distribution of four male and four female teams.

4.4.2. View Analysis

The videos of the last eight teams were analysed for view overlaps between participants, that is, a percentage of the total time of each round was determined in which both participants looked at the same artefact (quotes, photos, or Map-o-Mat). This analysis was performed to the same subset of groups that were used for the transcription analysis, since it was meant to complement and help interpret the findings of the transcription analysis.

For the view analysis, the HITLabNZ's in-house analysing tool "VideoAnalysisApp" [45] was used. This program allows the experimenter to open a video and press predefined buttons to record activity states of interest along a shared time line (see Fig. **6**).

For every condition, both participants' views of quotes, photos, or Map-o-Mat were recorded separately. Then, the summed time of view overlaps was calculated and divided by the total time, which produced the percentage of the view overlap in that condition.

4.5. Expected Results

The spatial embodied interaction afforded by the video-CVEs allows for non-verbal forms of communication such as gaze awareness and proximity behaviour which are not supported in sVC. We therefore predicted social presence of both video- CVEs to be higher than in sVC. We further pre-

[1] http://www.liwc.net, last accessed in July 2011.

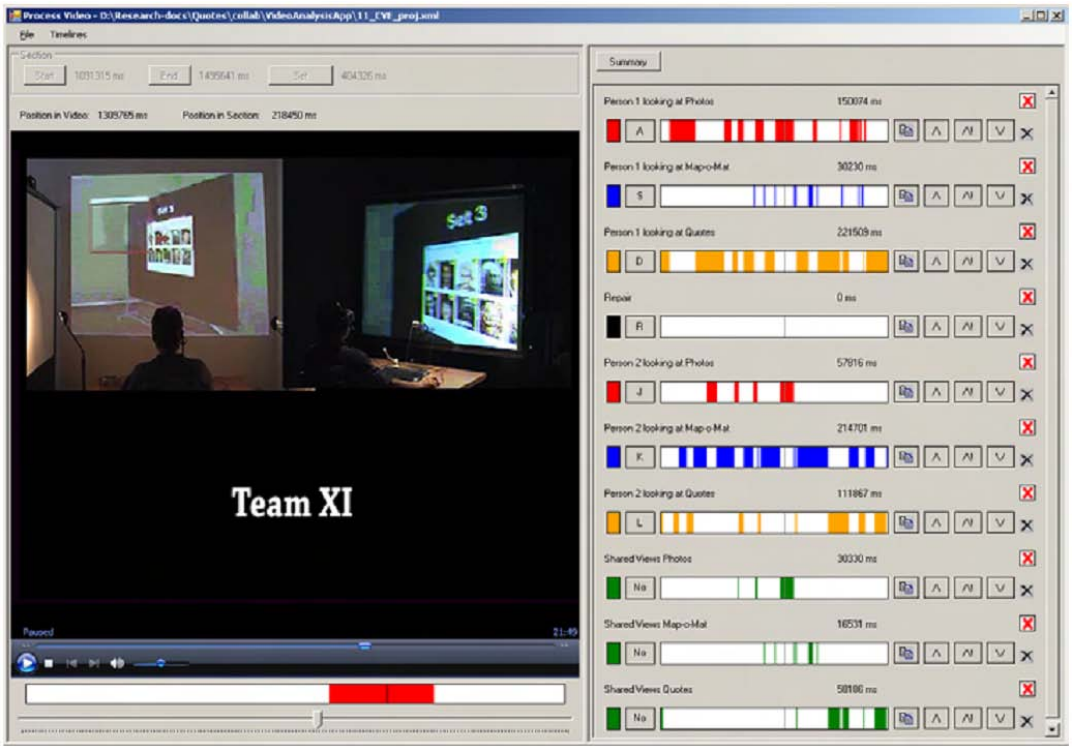


Fig. (6). The video analysis application. The bars on each time line on the right represent different views for each of the participants that can then be compared for overlaps. There is a view overlap in the snapshot depicted, since both participants are looking at the photos in that moment.

sumed that the stereo-projected video-CVE would induce a stronger sense of physical presence. Based on the correlation of physical presence and social presence reported by Nowak *et al.* (2001), we expected a higher level of physical presence to be paralleled by an increase in social presence in condition vCVE im. We also assumed that social presence in face-to-face would be highest.

Our presumptions for the speech analysis were that the awareness and attention benefits of the spatial embodied interaction in the video-CVEs would lead to more efficient verbal communication and referencing. In line with the findings of Kramer *et al.* [46], we also expected a higher sense of presence to lead to more face-to-face-like communication patterns.

Our final assumption was that participants would generally try to look at either photos or quotes at the same time in order to use easier forms of referencing (such as pointing and deictic references). Based on the different visual awareness mechanisms supported in each condition, we expected that the relative time both participants managed to look at the same object would be highest in face-to-face, and lowest in sVC.

5. RESULTS

The questionnaire data was analysed with SPSS version 14.0. The significance level was set to 0.05.

Data were analysed using Mixed Models ANOVA with "Medium" and "Gender" as fixed factor, and "Subject ID" nested within "Pairs" as random factor. It must be noted here that "Gender" was included as an additional between-subjects factor *ex post* in response to the salient differences we observed in the collaborative style of male and female teams during the experiment.

If a significant main effect was found between the four media conditions, pair-wise comparisons were performed using the Bonferroni adjustment for multiple comparisons. Further, Pearson correlations were calculated to explore the relationship between the different measures used.

5.1. Questionnaire Results

In total, 18 sessions with two participants each were run. These form the basis of the questionnaire results reported below. All questionnaires of the 36 subjects were valid, no values were missing.

5.1.1. Social Presence

The internal consistency of social presence based on the seven bipolar pairs was very good (Cronbach $\alpha = 0.91$).

There was a significant main effect ($F(3,48)=43$, $p<0.001$) between the four different media. Post-hoc comparisons showed that social presence was significantly higher in FtF ($M=6.1$, $SE=0.16$, $p<0.001$) than sVC ($M=4.4$, $SE=0.19$), vCVE_desk ($M=4.3$, $SE=0.17$), and vCVE_im ($M=4.4$, $SE=0.15$). Post-hoc comparisons showed significant differences between face-to-face and the three mediated conditions. No pair-wise differences were found between the three mediated conditions.

There was also a significant Medium X Gender interaction, $F(3,48)=5.7$, $p=0.002$. Of the mediated conditions, female participants rated sVC as highest in social presence and

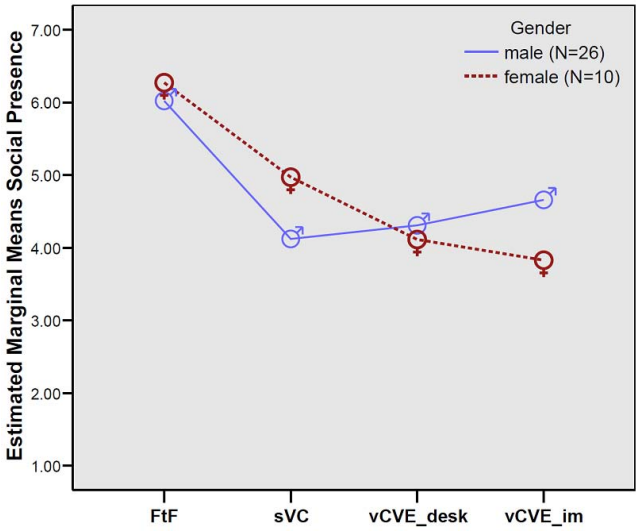


Fig. (7). Social presence by gender and medium.

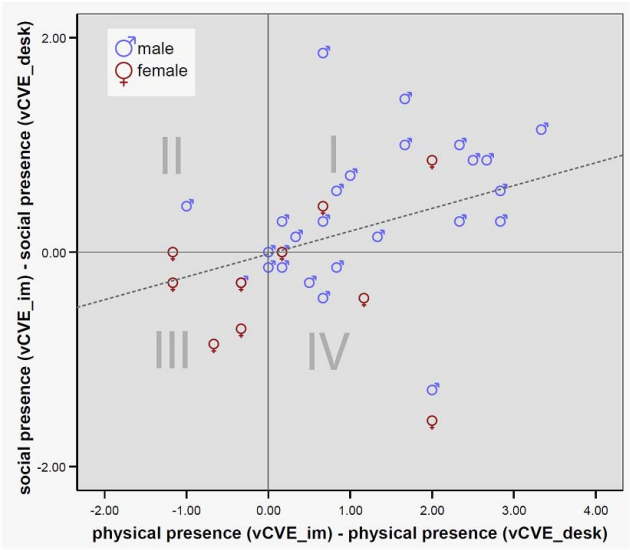


Fig. (8). Differences in physical presence and social presence between the two vCVE conditions, marked by gender.

vCVE_im as lowest, while for male participants the order was reversed (see Fig. **7**).

5.1.2. Preference

Among all participants, the condition FtF (M=1.3) was significantly preferred over sVC(M=2.7), vCVE_desk (M=2.9), and vCVE_im (M=3.1), (Friedman test, $\chi_r^2 = 40.8$, df=3, N=36, p<0.001).

Two more Friedman tests were carried out with the rankings of the three mediated conditions for male and female participants separately. No clear preference emerged for the male participants. However, female participants clearly preferred the standard videoconferencing interface (M=1.2) to the vCVE_desk interface (M=2.0) and vCVE_im (M=2.8), (Friedman test, $\chi_r^2 = 12.8$, df=2, N=10, p=0.002).

5.1.3. Physical Presence

The internal consistency of factor physical presence based on the six items was good (Cronbach α =0.89).

There was a significant effect between the two vCVE interfaces, F(1,34)=10, p=0.003, indicating a higher sense of physical presence was perceived in the more immersive vCVE condition.

There was also a significant interaction Medium x Gender, F(1,34)=4.6, p=0.039, suggesting an increase of physical presence in the more immersive vCVE condition was higher for male participants.

5.1.4. Social Presence and Physical Presence

We investigated whether the higher sense of physical presence induced by the more immersive video-CVE would lead to a higher perceived social presence of that video-CVE. Therefore, we first subtracted both physical presence and social presence scores obtained in condition vCVE_desk from those obtained in condition vCVE_im, and then combined the resulting difference scores in a scatter plot (see Fig. **8**) for visual inspection.

We also carried out a Pearson Correlation on the difference scores of physical presence and social presence. The test revealed positive correlation between the two dimensions, r=0.363, p=0.03, which Cohen (1988) calls a medium-sized effect. However, as can be seen by the number of data points in quadrant I, for only eighteen participants (50%) both physical and social presence increased in the more immersive video-CVE condition.

5.2. Linguistic Analysis

Because audio transcription is very time consuming, it was only possible to transcribe the conversations of the last eight teams (four male and four female teams). The Linguistic analysis and the following video-analysis of the view coordination are therefore only based on the data from the last 16 participants.

Approximately 23,000 words were transcribed from the audio files of the last eight teams by the first author using the open source software "Transcriber"[2].

The following rules were applied for the transcription process:

- Complete quotes that were read aloud by any participant were not transcribed word-by-word, but were marked as "Quote_Full" in the transcript.
- Verbal references to parts of quotes, like "the dogs in France one" or "...no eskimos in Iceland, hmmm, I think that could be Bjork" were marked as "Quote_Part, hmm, I think that could be Bjork" in the transcript.
- Verbal references to quote tags, such as "Do you think A6 could be Karl Marx?" were marked as "Do you think Quote_Tag could be Karl Marx?" in the transcript.
- Fillers, such as "you know", "I mean", and "I don't know" were transcribed as one word, "youknow", "Imean", and "Idontknow".

[2] http://trans.sourceforge.net/en/presentation.php

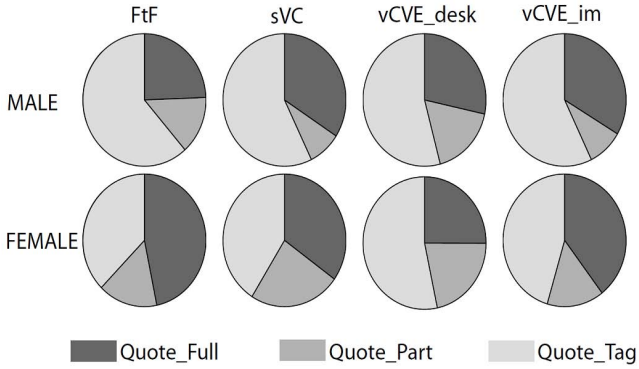


Fig. (9). References to quotes distinguished by QUOTES (whole quote, high verbal effort), QUOTE PARTS (only part of the quote is referenced, medium verbal effort), and QUOTE TAG (the quote is referenced by its tag, low verbal effort).

One separate text file was extracted for each participant and analysed with the software `LIWC2001´ [44] as mentioned above. The following passages present an overview of the most interesting findings:

5.2.1. LIWC Standard Dimensions

The transcription analysis revealed significant effects for the LIWC built in Linguistic categories "Space" and "Social Processes" (see Table **6**):

In all four conditions, female participants used more social expressions such as "talk", "us", or "friend" in their conversations, between-subjects effect (p=0.01).

There were also significant differences between the mediated conditions. Words with spatial character, such as "around", "over", or "up" occurred more frequently in both video-CVE conditions than in sVC or FtF (p<0.01).

5.2.2. Further Linguistic Dimensions

Three additional word categories were defined in a dictionary that was created specially for this experiment. These categories were laughter (e.g. "haha"), local deixis (e.g. this, here, these), and remote deixis (e.g. that, there, those). The relative occurrence of words belonging to these word categories were tested for effects.

There was a significant Medium X Gender interaction (p=0.07) for laughter. While female participants laughed the most in condition sVC and laughed least in the immersive vCVE condition, exactly the opposite was the case for male participants.

There was also an effect across all conditions with regard to the relative use of local deixis, according to which participants used local deixis most often in the FtF condition (p=0.07). No effect was found with regard to the occurrence of remote deixis.

5.2.3. References to Quotes as Measure of Verbal Effort

We were also interested in verbal references to quotes as a measure of verbal effort. Three different ways of referencing were compared:

- Reading out the whole quote to the other participant. This is verbally expensive, but does not require a shared contextual understanding.

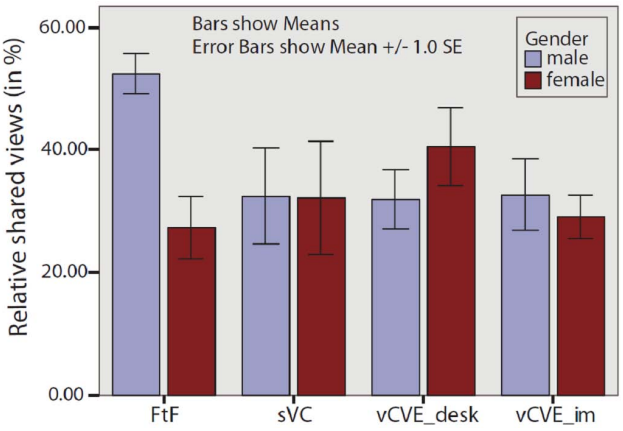


Fig. (10). Average of relative view overlap and standard error by gender and medium.

- Referring to a quote by reading or repeating only a significant part of it. This is verbally cheaper, but requires a certain level of shared contextual understanding.
- Referring to a quote only by the tag that was displayed on the side of each quote. This way of referencing involves the least verbal effort, but requires a high level of shared contextual information and interpersonal awareness.

The average of the relative occurrences of these three ways of referencing was determined for each condition. Fig. (**9**) shows a separate pie chart for every condition and gender.

As can be seen in Fig. (**9**), male participants generally used QUOTE TAG references more often than female participants (between-gender effect, F(1,14)=16, p=0.001). This difference was most obvious in condition FtF.

5.3. View Analysis

The videos of the last eight teams were analysed for view overlaps between participants. For every condition, both participants' views of either, quotes, photos, or Map-o-Mat were first marked as continuous events and then compared using our in-house video annotation and analysis tool [45]. Then, the relative time that both team members were looking at the same object was determined. The averages of the view overlaps are shown in Figure 10 for the four male and four female teams.

Male participants made bigger efforts to look at the same artefact as their team member, often resulting in one following the other around the table in order to maintain the shared view of what was lying in front of them and being talked about. Female participants, on the other hand, often positioned themselves at opposite ends of the table, where one had access to the photos and the other one to the quotes. They then talked about them while facing each other (see Fig. **10**).

5.4. Completion Time and Correct Answers

Also with regard to completion time and the number of correctly entered quotes, participants seemed to perform better in the Face-to-Face condition compared to the mediated

conditions, with an average completion time of 5.5 minutes in FtF compared to 7.5 minutes in sVC and 8.8 minutes in the CVE conditions, and with an average of 0.8 wrong answers being the lowest in FtF compared to 1.4, 1.6, and 1.2 wrong answers in sVC, vCVE_desk, and vCVE_im, respectively. It is however noted that the experiment task was designed with a clear focus on process and satisfaction measures and not on product measures,. The above mentioned results are therefore included here for completeness only.

6. DISCUSSION

The user experiments described in this paper assessed and compared aspects of user experience and collaborative behaviour afforded by face-to-face, two video-CVEs, and a standard VC condition. Our motivation was to learn about how video-CVEs influence collaboration, how we may improve them in the future, and which human factors we need to better understand. In the following sections we discuss our findings.

6.1. Does 'Being there' Help to 'Come Together'?

The condition vCVE im was rated marginally higher in social presence than condition vCVE desk by male participants, and was rated marginally lower in social presence than condition vCVE desk by female participants.

This suggests that the more immersive system created a more face-to-face-like user experience for male participants, while it created a less face-to-face-like user experience for female participants.

However, physical presence and social presence did not correlate as strongly as expected. The medium-sized positive correlation we found confirms the findings of Nowak [46]. In our case, we think that this correlation was negatively affected by the introduction of new disturbing factors in the immersive condition. For example, two participants reported light forms of motion sickness which had a negative impact on their social presence ratings. The technical overhead of condition vCVE im also sometimes caused discomfort. Participants (especially female participants) did not like to wear the shutter glasses, others were not used to the large field-of-view display.

The analysis of word counts could not replicate the presence correlations with the word categories found by Kramer *et al.* [47], probably because of the different scenario and task applied. However, the word categories "Space" and "Motion" showed some potential effects which may have been influenced by a sense of presence and thus may be candidates for a objective presence measure in scenarios that are more similar to the one applied in this experiment. This, however, needs to be investigated further. The occurrence of laughter showed a gender interaction in the two vCVE conditions and could be interpreted as a direct measure of enjoyment, suggesting that male participants enjoyed the immersive experience more.

In summary, the more immersive interface enhanced the user experience for participants who liked the immersive experience. For others, being immersed felt awkward, partly because the delivery of the more immersive experience came at the cost of a technical overhead that was too novel, too overwhelming, or too obtrusive. Because of the complexity of the presence phenomenon itself, more studies are needed to investigate the value of a sense of presence for remote collaboration in CVEs.

6.2. Gender Matters

Surprisingly, there were substantial differences in the observed collaboration styles of male and female teams in the unmediated FtF condition. Male friends made an effort to follow each other around the table in order to be able to look at the same artefact while talking to each other. The shared visual context allowed them to reduce their verbal effort, which resulted in the higher occurrence of local deixis and QUOTE TAG references. Female participants, in contrast, made less conscious efforts to share the same views when collaborating around the table and instead frequently placed themselves at the opposite ends of the table from where they would read the full quotes and describe the photos to each other verbally. They therefore accepted the need for a higher verbal effort for referencing in exchange for the ability look at the other person while talking to her. Female conversations also contained more words of the category "social processes" than male conversations.

These behaviours are in line with Wright's [48] observations according to which for "men friendship tends to be a side-by-side relationship, with the partners mutually oriented to some external task or activity; while for women friendship tends to be a face-to-face relationship, with the partners mutually oriented to a personalized knowledge of and concern for one another."

We believe that the different collaboration and communication styles we encountered for male and female teams are based on such fundamental differences in the way men and women communicate. We further believe that the different collaborative systems used in this study supported these collaborative styles more or less quite well.

The sVC condition allowed participants to see their partner's faces at all times, but did not provide visual awareness cues. This supported the inter-personal, verbal collaboration style adopted by female teams, who therefore rated both social presence and awareness higher in this condition. Compared to men, women use more facial expressions (Hall [49], page 71) and gaze at each other more, especially in same-sex dyads ([49], page 83). Consequently, for women, seeing the other person's facial expressions in a talking head video may be more important and thus conveys more social presence in the standard video-conferencing condition.

The vCVE conditions created a shared action space which came at the cost of not being able to see the other person's video at all times. This supported the task-focused, side-by-side collaboration style adopted by the male teams, leading to higher scores in social presence and awareness compared to the sVC condition. Adding visual awareness cues while compromising the view of the other's face, however, was not considered beneficial for female participants, leading to a decrease in social presence and awareness. This result is in line with a finding of Argyle *et al.* [50], who studied the effects of visibility on interaction in a dyad. They encountered "considerable sex differences" with females less

comfortable in situations where they could not see their counterparts.

As mentioned before, gender effects have not been reported and discussed much in other cross-media studies despite the fact that already deGreef *et al.* [20] suspected that "it is quite possible that women experience a higher level of social presence, considering the large differences in communication behaviour between men and women".

The results of our study re-confirmed their conclusion. One direct implication of our findings is therefore that future studies investigating (video-) mediated communication should pay more attention to gender-specific differences.

6.3. Design for 'Least Collaborative Effort', Not 'Least Verbal Effort'

In FtF, male friends made an effort to follow each other around the table in order to be able to look at the same artefact while talking to each other. The shared visual context allowed them to reduce their verbal effort, which resulted, in the more frequent occurrence of Quote Tag references. Female participants, in contrast, made less conscious efforts to share the same views when collaborating around the table and instead frequently placed themselves at the opposite ends of the table from where they would read the full quotes and describe the photos to each other verbally. They therefore accepted the need for a higher verbal effort for referencing in exchange for the ability to look at the other person while talking to them.

This teaches a valuable lesson to designers of synchronous groupware. Namely that striving for verbal efficiency might not always reach the most natural and most efficient form of collaboration - especially, as in our study, if social aspects of communication are compromised.

6.4. Reciprocal Awareness Problems

In contrast to what we expected, the linguistic analysis and the view analysis did not produce any noteworthy differences between the video-CVEs and sVC conditions, which suggests that the expected awareness advantage in our video-CVEs did not result in substantial benefits for the collaborative process.

While transcribing the conversations recorded in the video- CVE conditions, we noticed one particular problem that had a direct influence on the speech patterns we recorded in our video-CVE: verbal effort was wasted in situations where the speaker and listener looked at the same artefact, but the speaker was not aware of it. Consider the following example:

P1 is looking at the billboard with quotes, P2 joined him after having entered a pair in the Map-o-Mat. P1 does not see P2's avatar because he is behind him. P1: Do you think Clint said "I always play women I would date"?

P2: I don't know, could also be "A4" P1: ... oh? OK. Or maybe "A5"? What do you think of "A5"? P2: No idea. Has he ever been in France?

When P1 referenced the quote "I always play women I would date" at first, he chose the verbally most expensive way by reading it out completely. However, from the answer given by P2, he could infer that P2 must also be looking at the quotes in that moment, and therefore adopted the cheap referencing style from that moment on.

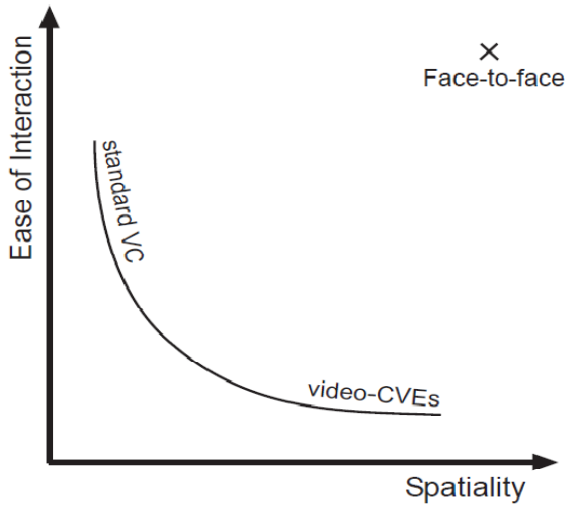


Fig. (11). The trade-off between supporting either the spatiality or the ease of interaction of face-to-face collaboration.

While being aware of speaker's context in a shared workspace helps listeners to interpret these utterances, it is crucial for the speaker to be aware of how aware their listeners are of this context in order to formulate verbal utterances in the most efficient way. While reciprocal awareness comes more naturally in Face-to-Face situations (people normally know what is going on around them and 'feel' if someone is looking over their shoulder), CVEs cannot provide the same level of peripheral awareness mainly because of a limited field of view and low fidelity, as well as a lack of other sensory stimuli such as smell or touch. Research in the area of CVEs should address this problem by further investigating how reciprocal awareness can be improved (see Fraser *et al.* [51] for suggestions).

6.5. Striving for the Gold Standard: a Trade-Off

Face-to-face was confirmed as the gold-standard for collaboration. It is highest social presence, communication efficiency and ease of use.

However, when attempting to support face-to-face-like tele-collaboration by the provision of spatial interfaces, one faces a dilemma: including spatial aspects may also introduce new interface techniques, especially for navigation,

which makes spatial systems harder to use.

Fig. (**11**) depicts the general underlying trade-off between spatiality and the ease of interactions.

Standard video-conferencing interfaces have a low level of spatiality, but are easier to use. Video-CVEs, in contrast, support spatiality and create a spatial context that is closer to face-to-face, but interactions come at the prize of a higher cognitive workload.

The value of spatial virtual video-conferencing should therefore be assessed in terms of a cost-benefit ratio which also takes advantages that can be gained into account. Fig. (**12**) shows a matrix representing this concept.

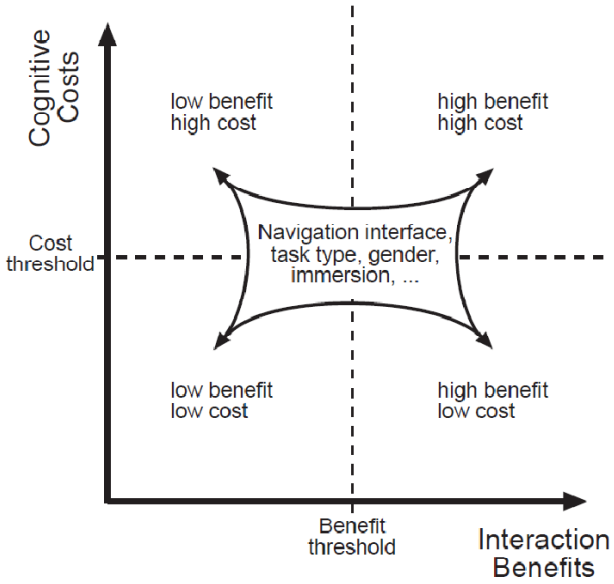


Fig. (12). The cost-benefit matrix for spatial tele-collaboration.

Cognitive costs are mainly due to navigation and object manipulation, but also include discomfort and distraction with complex hardware, confusion, the handling of a limited field-of-view, and a suspension of disbelief.

Interaction benefits include a higher social presence, a higher sense of copresence, the enjoyment of feeling immersed and transported to a virtual space, and the exploitation of non-verbal communication channels and awareness mechanisms.

- Low benefit, high cost: video-CVEs which are hard to use, but do not deliver additional support for the user's collaboration requirements do not improve regular video-conferencing.
- Low benefit, low cost: video-CVEs which are marginally harder to use than regular video-conferencing, but which deliver some additional support for the user's collaboration requirements, may improve regular video-conferencing.
- High benefit, high cost: if a video-CVE is hard to use, but delivers a substantial benefit for collaboration, users may be willing to accept the extra work they have to put in to get it.
- High benefit, low cost: video-CVEs that are marginally harder to use, but deliver a substantial benefit for collaboration outperform regular video-conferencing systems.

The ratio between costs and benefits differs for every user. Some people find it harder to navigate their avatars than others, some enjoy a feeling of physical presence, others feel distressed, and, as found for the different collaborative styles adopted by male and female pairs in our experiment, the benefits of spatial virtual interaction appear to support the collaborative requirements of men more than those of women and are therefore appreciated more by men.

In our experiment, the cost / benefit ratio was affected by the navigation interface provided, the nature of the task, the level of immersion, and by the gender of the participants. Yet there are likely to be more factors that play a role and these are worth investigating in future work.

7. CONCLUSION AND FUTURE WORK

In this paper we presented the results of a study comparing a 2D standard video-conferencing system with a 3D desktop video-based collaborative virtual environment (video-CVE), a more immersive 3D stereo-projected video-CVE, and a face-to-face condition. Participating teams consisted of two same-gender friends.

7.1. Substantive Findings

In the face-to-face condition, a considerable difference in the collaborative style between male and female teams emerged which may explain gender differences in the perception of the different 2D and 3D videoconferencing interfaces. Male participants rated the immersive video-CVE highest in social presence, followed by the desktop based system and the standard video-conferencing tool. This suggested that for them, an increase of spatiality and a sense of presence in a video-CVE contributed to a more face-to-face-like experience. Female participants, however, rated social presence and preference of these systems in completely the opposite order.

Collaborative behaviour, assessed by analysis of communication patterns and view coordination, differed between mediated and unmediated conditions, and between male and female teams, but was not found to be significantly influenced by the three different videoconferencing interfaces.

7.2. Methodological Contributions

A speech analysis of extracted communication transcripts was conducted based on the principles of Linguistic Inquiry and Word Count. Mediated and non-mediated communication patterns were characterised along several linguistic dimensions which may serve as a good reference for researchers wanting to use this relatively new method for assessing communication based on linguistic features in other cross-media studies.

Besides testing, developing, and exploring the measures that were applied in the experiment, Moreover, the task and scenario used in this experiment was designed to allow the assessment of verbal effort based on different types of referencing mechanisms that could be extracted from the communication transcripts.

7.3. Future Work

To further improve the value of spatial virtual conferencing by means of video-CVEs, research attempts should focus on investigating possibilities for reducing the cognitive costs involved, while gaining a better understanding of potential benefits. Some possible directions to pursue are discussed in the following sections.

7.3.1. Possible Directions for Human Factors Research

Research into video-CVEs is arguably still in its infancy, and this study is only one of the first of many steps that need to be followed until we actually understand the full potential

of this medium. There are several paths researchers could follow that would extend the findings presented in this paper.

7.3.1.1. Exploring Further Demographic Factors

The gender-effects observed in our study showed the importance of taking demographic variations into account when assessing the value of different telecommunication interfaces. Yet, there may be other factors that play an important role in the perception of the value of video-CVEs. Factors that could be worth investigating include users' immersive tendencies, users' technophobic biases, or users' level of extroversion, among others.

7.3.1.2. Varying the Group Size

Future research should also investigate the value of video-CVEs for remote collaboration between more than three participants. The bigger the group size, the harder it is for members to manage their interdependent actions.

It could therefore be expected that the benefit of supported gaze awareness will therefore be more appreciated.

7.3.2. Investigating Cues for Reciprocal Awareness

The restricted peripheral awareness of the CVEs used in the experiment prevented users from noticing other avatars that was immediately beside or behind them. This meant that users were not always aware that other avatars might be looking over their shoulder and other users could actually see the same thing in that moment. In these situations, the chance for more efficient grounding mechanisms that are based on the reciprocal awareness of sharing the same visual context were not exploited, and the speech patterns observed were not as efficient as they could have been. Researchers and designers of video-CVEs who try to allow users to benefit from better exploiting effective grounding mechanisms in these situations should therefore investigate explicit cues that make it clear to all participants whenever they share the same view within a video-CVE. To name two examples, people's videos could be superimposed on the screen of every user as soon as their perspectives overlap to a certain degree, or "conditional telepointers" could be provided in video-CVEs, which only appear if someone else is sharing the same perspective.

CONFLICT OF INTEREST

There is no conflict of interest. The German Academic Exchange Service, DAAD, financially supported the first author. Marratech provided a full license of their videoconferencing tool for this study. Daimler AG Research and Technology supported our work by allowing us to use their software framework RTB.

ACKNOWLEDGEMENTS

We would like to thank the five anonymous reviewers for their time and effort. We thank Carl Gutwin for his suggestions and comments which influenced the design and analysis of this study. Finally, thanks to all participants in this study.

REFERENCES

[1] J. Short, E. Williams, and B. Christie, "*The Social Psychology of Telecommunications"*. New York: Wiley, 1976.

[2] C. Gutwin and S. Greenberg, "The effects of workspace awareness support on the usability of real-time distributed groupware," *ACM Trans. Comput.-Hum. Interact.*, vol. 6, pp. 243-281, 1999.

[3] A. Chapanis, "Interactive human communication," *Sci. Am.*, vol. 232, pp. 36-42, 1975.

[4] W. R. Reitman, "*Cognition and Thought: An Information-Processing Approach"*, Wiley: NY" 1965.

[5] J. S. Olson, G. M. Olson, and D. K. Meader, "What mix of video and audio is useful for small groups doing remote real-time design work?," presented at the CHI '95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1995.

[6] C. O'Malley, S. Langton, A. Anderson, G. Doherty-Sneddon, and V. Bruce, "Comparison of face-to-face and video-mediated interaction," *Interact. Comput.*, vol. 8, pp. 177-192, 1996.

[7] G. Doherty-Sneddon, C. O'Malley, S. Garrod, A. Anderson, S. Langton, and V. Bruce, "face-to-face and video-mediated communication: a comparison of dialogue structure and task performance," *J. Exp. Psychol.: Appl.*, vol. 3, pp. 105-125, 1997.

[8] E. S. Veinott, J. Olson, G. M. Olson, and X. Fu, "Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other," presented at the CHI '99: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1999.

[9] R. D. Luce and H. Raiffa, "*Games and Decisions"*, Dover Publication: USA 1957.

[10] H. Wichman, "Effects of isolation and communication on Cooperation in a two-person game," *J. Per. Soc. Psychol.*, vol. 16, pp. 114-120, 1970.

[11] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, "Effects of four computer-mediated communications channels on trust development," presented at the CHI '02: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2002.

[12] J. A. Short, "Effects of Medium of Communication on Experimental negotiation," *Hum. Relat.*, vol. 27, pp. 225-234, 1974.

[13] A. Monk, J. McCarthy, L. Watts, and O. Daly-Jones, "Evaluation for CSCW," in *Measures of Process*, Springer: USA, 1996, pp. 125-139.

[14] B. O'Conaill, S. Whittaker, and S. Wilbur, "Conversations over Video conferences: an evaluation of the spoken aspects of video-mediated communication," *Hum-Comput. Interact.*, vol. 8, pp. 389-428, 1993.

[15] A. F. Monk and C. Gale, "A look is worth a thousand words: full gaze awareness in video-mediated conversation," *Discourse Processes*, vol. 33, pp. 257-278, 2002.

[16] A. J. Sellen, "Remote conversations: the effects of mediating talk with technology," *Hum.-Comput. Interact.*, vol. 10, pp. 401-444, 1995.

[17] E. Williams, "Experimental comparisons of face-to-face and mediated communication: a review," *Psychol. Bull.*, vol. 84, pp. 963-976, 1977.

[18] J. C. Tang, "*Why Do Users Like Video? Studies of Multimedia-Supported Collaboration"*, Technical Report 1992.

[19] J. Short, E. Williams, and B. Christie, "*The Social Psychology of Telecommunications"*, John Wiley: NY 1976.

[20] P. de Greef and W. A. IJsselsteijn, "Social Presence in a Home Tele-Application," *Cyberpsychol. Behav.*, vol. 4, pp. 307-315, 2001.

[21] J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn, "Spatiality in videoconferencing: trade-offs between efficiency and social presence," presented at the CSCW '06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, 2006.

[22] S. Whittaker, "Rethinking video as a technology for interpersonal communications: theory and design implications," *Hum.-Comput. Stud.*, vol. 42, pp. 501-529, 1995.

[23] S. Wheeler, "User reactions to videoconferencing: Which students cope best?," *Educ. Media. Int.*, vol. 37, pp. 31-38, 2000.

[24] H. Maurin, D. H. Sonnenwald, B. Cairns, J. E. Manning, E. B. Freid, and H. Fuchs, "Exploring gender differences in perceptions of 3D telepresence collaboration technology: an example from emergency medical care," presented at the Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles, Oslo, Norway, 2006.

[25] A. Stuhlmacher, M. Citera, and T. Willis, "Gender Differences in virtual negotiation: theory and research," *Sex Roles*, vol. 57, pp. 329-339, 2007.

[26] R. Wachter, "The effect of gender and communication mode on conflict resolution," *Comput. Hum. Behav.*, vol. 15, pp. 763-782, 1999.

[27] C. Teoh, H. Regenbrecht, and D. O'Hare, "Investigating Factors influencing Trust in Video-Mediated Communication", Proceedings of ACM OZCHI 2010, November 22-26, Brisbane, Australia, 2010.

[28] C. Teoh, H. Regenbrecht, and D. O'Hare, "The Transmission of Self: Body Language Availability and Gender in Videoconferencing", in Proceedings of ACM OZCHI 2011, Nov 28 - Dec 2, Canberra, Australia, 2011.

[29] D. Nguyen and J. Canny, "More than face-to-face: Empathy effects of video framing", in Proc. CHI 2009, ACM Press, pp. 423-432, 2009.

[30] R.J. Lowden and C. Hostetter, "Access, utility, imperfection: The impact of videoconferencing on perceptions of social presence", *Comput. Hum. Behav.*, Available online 1 November 2011, 10.1016/j.chb.2011.10.007 (http://www.sciencedirect.com/science-/article/pii/S0747563211002159)

[31] S. Benford, C. Greenhalgh, T. Rodden, and J. Pycock, "Collaborative virtual environments," *Commun. ACM*, vol. 44, pp. 79-85, 2001.

[32] S. Parise, S. Kiesler, L. Sproull, and K. Waters, "My partner is a real dog: cooperation with social agents," presented at the CSCW '96: Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, 1996.

[33] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperat. Virtual Environ.*, vol. 12, pp. 481-494, 2003.

[34] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," presented at the CHI '03: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2003.

[35] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich, "The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments," *Presence: Teleoperat Virtual Environ*, vol. 14, pp. 379-393, 2005.

[36] C. Greenhalgh and S. Benford, "MASSIVE: a collaborative virtual environment for teleconferencing," *ACM Trans. Comput.-Hum. Interact.*, vol. 2, pp. 239-261, 1995.

[37] V. Normand, C. Babski, S. Benford, A. Bullock, S. Carion, Y. Chrysanthou, N. Farcet, E. Frecon, J. Harvey, N. Kuijpers, N. Magnenat-Thalmann, S. Raupp-Musse, T. Rodden, M. Slater, G. Smith, A. Steed, D. Thalmann, J. Tromp, M. Usoh, G. V. Liempd, and N. Kladias, "The COVEN Project: Exploring Applicative, Technical, and Usage Dimensions of Collaborative Virtual Environments," *Presence: Teleoperat. Virtual Environ.*, vol. 8, pp. 218-236, 1999.

[38] J. Bowers, J. Pycock, and J. O'Brien, "Talk and embodiment in collaborative virtual environments," presented at the CHI '96: Proceedings of the SIGCHI conference on Human factors in Computing Systems, 1996.

[39] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder, "Small-Group Behaviour in a Virtual and Real Environment," *Presence: Teleoperat. Virtual Environ.*, vol. 9, pp. 37-51, 2000.

[40] H. Nakanishi, C. Yoshida, T. Nishimura, and T. Ishida, "FreeWalk: A Three-Dimensional Meeting-Place for Communities," in *Community Computing: Collaboration over Global Information Networks*, John Wiley and Sons, pp. 55-89, 1998.

[41] E.-L. Sallnäs, "Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments," *Presence: Teleoperat. Virtual Environ.*, vol. 14, pp. 434-449, 2005.

[42] H. Regenbrecht, T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller, "Using augmented virtuality for remote collaboration," *Presence: Teleoperat. Virtual Environ.*, vol. 13, pp. 338-354, 2004.

[43] T. Schubert, F. Friedmann, and H. Regenbrecht, "The Experience of Presence: Factor Analytic Insights," *Presence: Teleoperat. Virtual Environ.*, vol. 10, pp. 266-281, 2001.

[44] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count: LIWC2001," 2003.

[45] J. Looser, "Analysing Videos with VideoAnalysisApp: Manual," HITLabNZ, 2007.

[46] K. Nowak, "Defining and Differentiating Copresence, Social Presence and Presence as Transportation," presented at the PRESENCE 2001 - 4th Annual International Workshop, 2001.

[47] A. D. I. Kramer, L. M. Oh, and S. R. Fussell, "Using linguistic features to measure presence in computer-mediated communication," presented at the CHI '06: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2006.

[48] P. H. Wright, "Men's friendships, women's friendships and the alleged inferiority of the latter," *Sex Roles*, vol. 8, pp. 1-20, 1982.

[49] J. A. Hall, "*Nonverbal Sex Differences*," Johns Hopkins University Press: USA 1990.

[50] M. Argyle, M. Lalljee, and M. Cook, "The Effects of Visibility on Interaction in a Dyad," *Hum. Relat.*, vol. 21, pp. 3-17, 1968.

[51] M. Fraser, S. Benford, J. Hindmarsh, and C. Heath, "Supporting awareness and interaction through collaborative virtual interfaces," presented at the UIST '99: Proceedings of the 12th annual ACM Symposium on User Interface Software and Technology, 1999.

Received: February 02, 2012 Revised: February 16, 2012 Accepted: March 11, 2012