Understanding Remote Collaboration in Video Collaborative Virtual Environments

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Doctor of Philosophy
in the
University of Canterbury
by

Jörg Hauber

Supervising Committee

Dr. Andy Cockburn	Supervisor
Dr. Mark Billinghurst	Co-Supervisor
Dr. Holger Regenbrecht	Associate Supervisor
Dr. Dean Owen	Associate Supervisor

University of Canterbury
2008

Publications from this dissertation

Material from this dissertation has been previously published in the peer-reviewed papers, posters, and extended abstracts listed below. The chapters of this thesis that relate to each publication are noted. I also comment on my individual contribution to each of these papers in the alongside paragraphs.

1. Hauber, J., Regenbrecht, H., Hills, A., Cockburn, A. and Billinghurst, M. (2005) Social Presence in Two- and Three-dimensional Videoconferencing. Presented at PRESENCE 2005, 8th Annual International Workshop on Presence, (London, Sep. 21-23, 2005). (Chapter 4)

2. Hills, A., Hauber, J. and Regenbrecht, H. (2005) Videos in Space: A study on Presence in Video Mediating Communication Systems. Presented at the 15th International Conference on Artificial Reality and Telexistence (ICAT 2005), (Christchurch, Dec. 5-8, 2005). (Chapter 5)

3. Regenbrecht, H., Haller, M., Hauber, J., Billinghurst, M. (2006) Carpeno: Interfacing Remote Collaborative Virtual Environments with Table-Top Interaction. Virtual Reality - Systems, Development and Applications, 10(2) (September 2006). Springer/London, 95–107. (Chapter 6)

4. Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. (2006) Spatiality in videoconferencing: trade-offs between efficiency and social presence. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (Banff, Alberta, Canada, November 04-08, 2006). CSCW '06. ACM Press, New York, NY, 413–422. (Chapter 6)

5. Hauber, J. (2006) Improving Video-Mediated Communication with a Collaborative Virtual Environment Approach. Doctoral Colloquium Abstract, ACM Conference on Computer Supported Co-operative work (CSCW '06), Banff, Alberta, Canada, November 2006.

6. Hauber, J., Regenbrecht, H., Billinghurst, M., Teoh, H. (2006) Presence in Video-mediated Communication: Does being there help to come together? Poster, ACM Conference on Computer Supported Co-operative work (CSCW '06), Banff, Alberta, Canada, November 2006. (Chapter 7)

Most of the papers resulted from collaborative work. My contributions to these papers were (numbers correspond to list numbers above):

1. This study was run by me with the assistance of Aimee Hills and under the supervision of Dr. Regenbrecht. I was responsible for experiment design and data analysis. I also wrote the paper with help from Dr. Regenbrecht, Dr. Cockburn, and Dr. Billinghurst.

2. This study was designed and conducted collaboratively by Aimee Hills and me, again under supervision and assistance of Dr. Regenbrecht. However, I analysed the gathered data separately and wrote most of the paper.

3. The Carpeno project was primarily a research effort by Dr. Regenbrecht and Dr. Haller. I helped with the implementation of the prototype and conducted two informal user studies during its demonstration at two local conferences. I also wrote the part of the paper describing findings of these studies. The experiences gathered with both hardware and user feedback influenced the design of my follow-up experiments.

4. This was my own research. My supervisors Dr. Regenbrecht, Dr. Billinghurst, and Dr. Cockburn consulted me during the design and data analysis phase and repeatedly reviewed the resulting paper I wrote.

5. This extended abstract was written by me with assistance of my supervisors Dr. Cockburn and Dr. Billinghurst.

6. This poster was designed by me and reflected my ideas for the design of a study that I carried out later in a slightly modified form (see Chapter 7).

Signature Date

Dedicated to Sandra

Abstract

Video-mediated communication (VMC) is currently the prevalent mode of telecommunication for applications such as remote collaboration, teleconferencing, and distance learning. It is generally assumed that transmitting real-time talking-head videos of participants in addition to their audio is beneficial and desirable, enabling remote conferencing to feel almost the same as face-to-face collaboration. However, compared to being face-to-face, VMC still feels distant, artificial, cumbersome, and detached. One limitation of standard video-collaboration that contributes to this feeling is that the 3D context between people and their shared workspace given in face-to-face collaboration is lost. It is therefore not possible for participants to tell from the video what others are looking at, what they are working on, or who they are talking to.

Video Collaborative Virtual Environments (video-CVEs) are novel VMC interfaces which address these problems by re-introducing a virtual 3D context into which distant users are mentally "transported" to be together and interact with the environment and with each other, represented by their spatially controllable video-avatars. To date, research efforts following this approach have primarily focused on the demonstration of working prototypes. However, maturation of these systems requires a deeper understanding of human factors that emerge during mediated collaborative processes.

This thesis contributes to a deeper understanding of human factors. It investigates the hypothesis that video-CVEs can effectively support face-to-face aspects of collaboration which are absent in standard video-collaboration. This hypothesis is tested in four related comparative user studies involving teams of participants collaborating in video-CVEs, through standard video-conferencing systems, and being face-to-face. The experiments apply and extend methods from the research fields of human-computer interaction, computer-supported cooperative work, and presence.

Empirical findings indicate benefits of video-CVEs for user experience dimensions such as social presence and copresence, but also highlight challenges for awareness and usability that need to be overcome to unlock the full potential of this type of interface.

Table of Contents

List of Figures **viii**

List of Tables **xi**

Chapter 1: Introduction **1**

1.1 Research context 2
1.2 Problem statement and research hypothesis 2
1.3 Contributions 4
1.3.1 Methodological contributions 4
1.3.2 Substantive contributions 4
1.4 Structure of this thesis 5

Chapter 2: Background **8**

2.1 Getting a grip on communication 9
2.1.1 Shannon and Weaver's model of communication 9
2.1.2 Interpersonal communication 12
2.1.3 Collaboration 17
2.1.4 Criticisms of transmission models 20
2.1.5 Summary 21
2.2 Mediated communication 22
2.2.1 Media characteristics 22
2.2.2 Media selection 30
2.2.3 Remote collaboration in shared workspaces 32
2.3 Video-mediated communication 35
2.3.1 Review of empirical work in video-mediated communication 36
2.3.2 Improving video-mediated communication 47
2.4 Using virtual reality (VR) as a communication medium 57
2.4.1 The sense of presence in virtual environments 58

2.4.2 Collaborative virtual environments (CVE) 63
2.4.3 The affordances of CVEs for remote collaboration . . . 64
2.4.4 Examples of CVEs 67
2.4.5 Empirical research in CVEs 70
2.5 Chapter summary 73

Chapter 3: Research Approach 75
3.1 A CVE approach to video-conferencing 75
3.1.1 The video-CVE "cAR/PE!" 77
3.2 Method 82
3.2.1 A comparative approach 83
3.2.2 Independent measures 84
3.2.3 Dependent measures 87
3.3 Overview of the experiments 88

Chapter 4: Experiment: "Desert Survival Game" 93
4.1 Experiment design 95
4.1.1 Experimental task 95
4.1.2 Experiment conditions 96
4.1.3 Apparatus 97
4.1.4 Participants 97
4.1.5 Procedure 98
4.1.6 Measures 100
4.2 Data analysis 100
4.2.1 Social presence as quality of the medium 100
4.2.2 Social presence as the experience of the other 101
4.2.3 Exploratory questionnaire items 101
4.2.4 Scale validity 102
4.2.5 Age and simulator experience effects 103
4.3 Results 103
4.3.1 Social presence as a quality of the medium 103
4.3.2 Social presence as the experience of the other 104
4.3.3 Exploratory questionnaire items 105
4.3.4 Scale validity 106

4.3.5 Age and simulator experience effects 108
4.3.6 User observation in condition vCVE 108
4.4 Discussion 110
4.5 Experiment summary 111

Chapter 5: Experiment: "Dream House" 114
5.1 Experiment design 115
5.1.1 Experimental task 115
5.1.2 Experiment conditions 116
5.1.3 Apparatus 117
5.1.4 Participants 118
5.1.5 Procedure 119
5.1.6 Measures 119
5.2 Data analysis 120
5.2.1 Social presence 120
5.2.2 Physical presence 121
5.2.3 The relation between physical presence and social presence 121
5.2.4 Reliability analysis 121
5.2.5 Preference ratings 121
5.2.6 Exploration of other factors 121
5.3 Results 122
5.3.1 Differences in social presence 122
5.3.2 Differences in physical presence 122
5.3.3 Exploring the relationship between social presence and physical presence 122
5.3.4 Reliability analysis 124
5.3.5 User comments and preference 126
5.3.6 Gender and relationship effects 127
5.4 Discussion 127
5.5 Experiment summary 128

Chapter 6: Experiment: "Dogs and Owners" 130
6.1 Experiment design 132

6.1.1 Experimental task . . . 132
6.1.2 Experiment conditions . . . 133
6.1.3 Apparatus . . . 135
6.1.4 Participants . . . 137
6.1.5 Procedure . . . 138
6.2 Measures . . . 139
6.2.1 Social presence . . . 139
6.2.2 Copresence . . . 139
6.2.3 Usability parameters . . . 140
6.2.4 Video recordings . . . 140
6.3 Data analysis . . . 140
6.3.1 Social presence . . . 140
6.3.2 Copresence . . . 140
6.3.3 Usability parameters . . . 141
6.3.4 Video analysis . . . 141
6.3.5 Preference . . . 141
6.3.6 Scale validity . . . 141
6.3.7 Gender effects . . . 141
6.4 Results . . . 142
6.4.1 Social presence . . . 142
6.4.2 Copresence . . . 143
6.4.3 Usability parameters . . . 143
6.4.4 Preference . . . 145
6.4.5 Video analysis . . . 146
6.4.6 Scale validity . . . 148
6.4.7 Gender effects . . . 148
6.5 Discussion . . . 149
6.6 Experiment summary . . . 150

Chapter 7: Experiment: "Celebrity Quotes" 152

7.1 Experiment design . . . 153
7.1.1 Experiment task . . . 154
7.1.2 Experiment scenario . . . 156
7.1.3 Experiment conditions . . . 157

7.1.4 Apparatus 159
7.1.5 Participants 161
7.1.6 Procedure 161
7.2 Measures 163
7.2.1 Social presence 163
7.2.2 Physical presence 163
7.2.3 Items on awareness, copresence, and ease of use 163
7.2.4 Video recordings 164
7.2.5 Communication transcripts 164
7.3 Data analysis 165
7.3.1 Social presence 165
7.3.2 Physical presence 165
7.3.3 Copresence, awareness, and ease of use 166
7.3.4 Preference 166
7.3.5 Linguistic features analysis 166
7.3.6 View analysis 167
7.4 Results 167
7.4.1 Questionnaire results 167
7.4.2 Results preference 174
7.4.3 Transcription analysis 175
7.4.4 Video analysis 184
7.5 Discussion 185
7.5.1 The impact of the "remoteness" of a video-CVE 185
7.5.2 Gender effects 187
7.5.3 Transcription analysis 188
7.5.4 Reciprocal awareness issues in the vCVE conditions . . 189
7.6 Experiment summary 190

Chapter 8: Discussion 192

8.1 User experience of spatial virtual video-conferencing 193
8.1.1 Gender effect 194
8.1.2 Spatiality and the feeling of being face-to-face 195
8.1.3 Transportation and the feeling of being face-to-face . . 195
8.2 User behaviour in spatial virtual video-conferencing 196

8.3 Striving for the gold standard: a trade-off 197
8.4 Future work 200
8.4.1 Possible directions for human factors research 200
8.4.2 Possible directions for video-CVE interface design . . . 202
8.5 Critical reflection 204
8.5.1 Reliability 205
8.5.2 Construct validity 206
8.5.3 External validity 206
8.5.4 Internal validity 208

Chapter 9: Conclusion 209
9.1 Summary of experiments 210
9.1.1 Experiment "Desert Survival Game" 210
9.1.2 Experiment "Dream House" 211
9.1.3 Experiment "Dogs & Owners" 211
9.1.4 Experiment "Celebrity Quotes" 212
9.2 Key questions addressed in this work 213
9.3 Main contributions 215
9.3.1 Methodological contributions 215
9.3.2 Substantive contributions 216
9.3.3 Minor contributions 217
9.4 Conclusion 217

References 219

Appendix A: Questionnaires 242
A.1 Desert Survival Game 242
A.2 Dream House 249
A.3 Dogs & Owners 253
A.4 Celebrity Quotes 257

Appendix B: Factor Analysis 260
B.1 Experiment "Desert Survival Game": Exploratory questionnaire items. 260

Appendix C: Transcription analysis **262**
C.1 Transcription excerpt 262
C.2 LIWC2001 Output Variable Information 266

List of Figures

1.1 Two people collaborating in three different media conditions. . . 3

2.1 Shannon and Weaver's model of communication. 10
2.2 The Shannon-Weaver model of interpersonal communication (simplified from DeVito (1998)). 13
2.3 The Lasswell formula (from Lasswell 1948). 15
2.4 A simple model for collaboration. 18
2.5 Richness of different communication media (from Daft et al. (1987)). 23
2.6 Example of a bipolar scale. 26
2.7 Social presence of several communication media (from Short et al. (1976)). 27
2.8 The Networked Minds definition of social presence (from Biocca et al. (2001)) 29
2.9 A general framework of presence (from IJsselsteijn and Riva (2003)). 59
2.10 The relationship between physical presence and social presence (from IJsselsteijn et al. (2001)). 61
2.11 An example of a CVE. 65

3.1 Examples of avatars in online 3D worlds. 76
3.2 Video avatars in Vertegaal et al.'s "Gaze-2" 76
3.3 The cAR/PE! virtual conferencing room. 77
3.4 A video-avatar in cAR/PE!. 78
3.5 The Mona Lisa effect. 79
3.6 Examples for gaze awareness in cAR/PE!. 80
3.7 The cAR/PE! network architecture (adapted from Regenbrecht et al. (2006) 82
3.8 Spatiality of shared spaces and examples. 85

3.9 Transportation in shared spaces and examples. 86
3.10 Artificiality of shared spaces and examples. 87

4.1 Conditions in the experiment "Desert Survival Game". 96
4.2 A video-conferencing workspace. 99
4.3 Results social presence as quality of the medium. 103
4.4 Results social presence as experience of the other. 104
4.5 Average and standard errors of extracted factors. 106
4.6 A frequently observed avatar navigation sequence. 109

5.1 The condition vCVE in the experiment "Dream House". . . . 116
5.2 The condition sVC in the experiment "Dream House". 117
5.3 A participant sitting at one of the workstations. 118
5.4 Differences in physical and social presence. 123
5.5 Differences in physical and social presence by relationship. . . 125

6.1 An example set of dogs and owners. 133
6.2 Conditions in the experiment "Dogs & Owners". 134
6.3 Social presence averages and standard errors. 143
6.4 Copresence average ratings and standard errors. 144
6.5 Interaction between conditions and gender for social presence. 149

7.1 An example set of quotes and photos. 154
7.2 The "Map-o-Mat" mapping application. 155
7.3 The virtual room used in the experiment "Celebrity Quotes". . 157
7.4 Conditions in the experiment "Celebrity Quotes". 158
7.5 The video analysis application. 168
7.6 Social presence by gender and medium. 169
7.7 Physical presence by gender and medium. 170
7.8 Physical presence and social presence in two vCVE conditions. 171
7.9 Awareness by gender and medium. 172
7.10 "Ease of use" by gender and medium. 173
7.11 Standard linguistic dimensions that showed significant effects. 177
7.12 Linguistic dimensions from the category "Psychol. Processes". 179
7.13 Linguistic dimensions from the category "Relativity". 180
7.14 Additional linguistic dimensions. 181

7.15 References to quotes. 183
7.16 View coordination. 184
7.17 Typical positioning of team members in condition FtF. 185

8.1 Social presence, averaged over all experiments. 193
8.2 The trade-off between spatiality and the ease of interactions. . 198
8.3 The cost-benefit matrix for spatial tele-collaboration. 199

List of Tables

2.1 Benefits of visual information for grounding subtasks (from Fussell et al. (2000)). 20
2.2 Constraints of different media for grounding processes (from Olson and Olson (2000)). 25
2.3 Example items of the NWM measure of social presence. 30
2.4 Grounding costs and examples (from Clark and Brennan (1991). 32
2.5 Elements of workspace awareness (from Gutwin (1997)). . . . 34
2.6 Cross-media studies by task and conditions compared. 37
2.7 Cross-media studies: product measures and results. 39
2.8 Cross-media studies: process measures and results. 42
2.9 Cross-media studies: satisfaction measures and results. 45
2.10 User studies in CVE-research. 71

3.1 The four experiments in chronological order. 88
3.2 Types of measures applied in the four studies 89
3.3 The experimental conditions and their level of transportation. 90
3.4 The experimental conditions and their level of spatiality. . . . 90
3.5 Overview of participants recruited in the four experiments. . . 92

4.1 Results social presence as experience of the other. 105
4.2 Factor analysis of exploratory questionnaire items. 105
4.3 Results of exploratory questionnaire items. 106
4.4 Factor loadings for social presence as quality of the medium. . 107
4.5 Reliability of the NWM measure of social presence. 108

5.1 Factor analysis of social presence. 126

6.1 The conditions in the experiment "Dogs & Owners". 133
6.2 Results of eleven usability items. 145
6.3 Results of the video analysis parameters. 146

7.1 Effects in category "Standard Linguistic Dimensions" 176
7.2 Effects in the category "Psychological Processes" 178
7.3 Effects in category "Relativity" 179
7.4 Effects in additional word categories. 180

C.1 LIWC2001 test dimensions and examples 267

Acknowledgements

Many people supported me during the last few years and I want to acknowledge them here.

A special thanks must go to my supervisors: Andy Cockburn for his patience and invaluable guidance during our weekly meetings; Mark Billinghurst for his belief in my abilities, and for always having been available when it really counted; Holger Regenbrecht for the fruitful collaboration with the University of Otago, and for the numerous long conversations we had that reached well beyond work-related topics; and finally Dean Owen for introducing me to the world of psychology.

Thank you also to all staff and students at the HIT Lab NZ who helped me out in so many ways and who I enjoyed sharing so many joyful moments with.

I would also like to take this opportunity to thank Aimee Hills, Gordon Wetzstein, Manuela Waldner, Daniel Wickeroth, Daniel Kerse, Anne Williams, Hon Leng Teoh, and Mira Hattal who helped run the experiments and contributed their time and effort.

Finally, I am grateful to all the encouragement and support I have received from friends and family, above all from my girlfriend Sandra who went through all this with me.

Thank you all!

Furthermore, I would like to acknowledge DaimlerChrysler Research and Technology, and Marratech for providing access to their video-conferencing software and applications.

This research was financially supported by a University of Canterbury Doctoral Scholarship, and by a doctoral scholarship of the German Academic Exchange Service (DAAD).

Abbreviations

ANOVA	analysis of variance
AO	audio-only
AV	audio-video
cAR/PE!	collaborative Augmented Reality for Presentation and Engineering
CAVE	cave automatic virtual environment
CSCW	computer supported cooperative work
CVE	collaborative virtual environment
DSG	desert survival game
FtF	face-to-face
HCI	human computer interaction
HMD	head-mounted display
IPQ	I-Group Presence Questionnaire
LIWC	Linguistic Inquiry and Word Count
LBE	Location-Based Entertainment
MMOG	massively multiplayer online game
MRA	mixed reality architecture
MUDs	Multi-User Dungeons
NWM	Networked Minds
PD	prisoner's dilemma
sVC	standard video-conferencing
vCVE	video-CVE
vCVE_desk	desktop-based video-CVE
vCVE_im	immersive video-CVE
VE	virtual environment
VMC	video-mediated communication
VR	virtual reality
vTab	video table
WYSIWIS	what you see is what I see

Chapter 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. Modern video-conferencing technology promises to deliver such an illusion, often heralding video-mediated communication (VMC) to be the next best thing to being face-to-face. This claim seems not too far-fetched if one considers the advantage of being able to see the facial expressions of the other person during video conversations which are absent in normal telephone calls. Compared to being face-to-face, however, even VMC still feels distant, artificial, cumbersome, and detached.

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

Video Collaborative Virtual Environments (video-CVEs) are novel VMC interfaces which seek to address these problems by re-introducing a virtual 3D context into which distant users are mentally "transported" to be together and interact with the environment and with each other. Although working prototypes of video-CVEs have demonstrated their technical feasibility, research into the value of video-CVEs to support remote collaboration is still in its infancy, and the human factors involved are not well understood.

This thesis addresses this lack of understanding by investigating, from the user's perspective, whether video-CVEs can effectively support otherwise missing face-to-face aspects of collaboration.

1.1 Research context

This research aligns with the field of Computer Supported Cooperative Work (CSCW), an area in the study of Human-Computer Interaction (HCI) that "examines how people work together in groups and how groupware technologies can support collaboration" (Ishii et al. 1994).

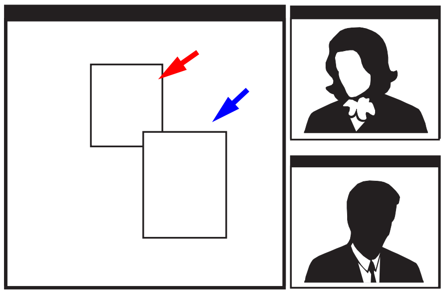
CSCW emerged as an identifiable inter-disciplinary research field in the early 1980s, a time in which networked personal workstations could be first used to communicate and work with others. In order to tackle the complex challenges posed by real-time distributed collaboration, CSCW advocates a "user-centred" design philosophy, in which the development and the evolution of interactive systems is tightly coupled with systematic evaluation and user-tests:

> "... Trial and error from creative system builders is too slow a discovery process. What is required is a better understanding of group work, the extent of the possibilities of the design space of technology features, and evaluation of systems in use that leads to a theory of computer supported cooperative work, which in turn can help us direct subsequent inventions of new ways to do group work." (Olson et al. 1993).

One area of CSCW focuses on real-time communication through video and audio. Researchers studied the value of different media conditions for remote collaboration (e.g. Chapanis 1975, Olson and Olson 2000), identified limitations of conventional video-conferencing setups (e.g. Gaver 1992, Whittaker 1995), and explored new interface approaches to improve VMC (e.g. Sellen 1992, Ishii et al. 1993, Nguyen and Canny 2005). This thesis extends this area by investigating the opportunities and challenges posed by the interface approach of video-CVEs.

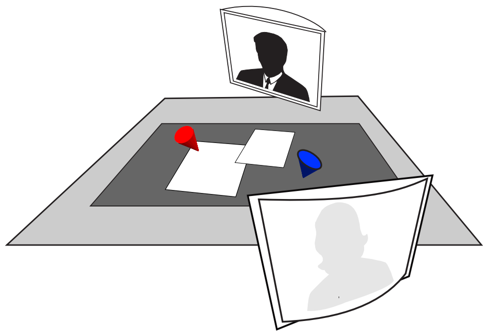
1.2 Problem statement and research hypothesis

The problem that this research addresses is that collaborating via standard video-conferencing systems feels artificial, cold, impersonal, and cumbersome



(a) Collaboration through a standard video-conferencing system

(b) Face-to-Face collaboration



(c) Collaboration in a video-CVE

Figure 1.1: Two people collaborating in three different media conditions.

compared to face-to-face interaction[1].

One cause of this problem is that standard video-conferencing systems do not provide the 3D frame of reference that we normally take for granted in face-to-face collaboration (see Figure 1.1(b)), and which affords a range of natural non-verbal communication cues including pointing in space, gaze direction, or proximity behaviour.

Video Collaborative Virtual Environments (video-CVEs) are novel video-conferencing interfaces which integrate spatially controllable video planes (video-avatars) of participants in a 3D virtual environment where they assume individual positions and viewing angles (see Figure 1.1(c)). Video-

[1] Within this thesis, "standard video-conferencing systems" refer to VMC systems, like that depicted in Figure 1.1(a), which provide "talking-head" videos of participants together with audio for communication next to a shared workspace area.

CVEs are inherently spatial and therefore afford forms of interaction which more closely resemble spatial face-to-face interaction.

The derived main hypothesis of this thesis is that **geographically dispersed people who collaborate in video-CVEs feel and behave more similarly to being face-to-face than if they collaborate through standard video-conferencing systems.**

This hypothesis is investigated and tested in four related comparative user studies involving participants collaborating in video-CVEs, through standard video-conferencing systems, and while being face-to-face. The studies apply process and satisfaction measures to assess and compare user experience factors and collaborative behaviour between these conditions.

The main goal of this research is to enhance our understanding of the opportunities and challenges posed by video-CVEs as a new communications medium.

1.3 Contributions

This thesis contributes substantive and methodological knowledge to the field of CSCW.

1.3.1 Methodological contributions

The methodological contributions made by this thesis concern the development and maturation of methods in the study of VMC.

1. Exploration and advancement of measures to discriminate user experience dimensions and collaboration patterns between variants of VMC interfaces.

2. Three collaborative tasks for use when studying remote collaboration.

1.3.2 Substantive contributions

The substantive contributions made by this thesis concern research findings of the impact of *spatiality* and *transportation* on user experience and collaboration. Spatiality and transportation are two inherent interface characteristics (Benford et al. 1996) that set video-CVEs apart from standard video-conferencing systems. Spatiality concerns an interface's level of support for fundamental physical spatial properties such as containment, topology, distance, orientation, and movement. Transportation concerns the degree to which an interface mentally "transports" its users into some new remote space in order to meet with others.

1. Research findings addressing the *spatiality* aspect of video-CVEs reveal improvements for the user experience such as an increase of social presence, but also uncover usability issues that still need to be overcome to reach the ease and efficiency of face-to-face collaboration.

2. Research findings addressing the *transportation* aspect of video-CVEs show only a weak correlation between the level of transportation of a video-CVE and the face-to-face-like experience and behaviour it affords.

3. Further findings include the detection of a gender effect, according to which video-CVEs are of more value for male users.

1.4 Structure of this thesis

Chapter 2 is about communication. It lays the theoretical foundations and reviews previous work for this research. First, it explains models and theories for interpersonal communication and collaboration that help us understand the challenges we face when mediating collaboration by means of technology. To foster this understanding, several communication media are also reviewed and characterised according to existing classifications. Then, it concentrates on the mediation of collaboration by means of video-conferencing technology. It gives an overview of cross-media studies involving audio, video, and face-to-face collaboration, and reviews different approaches to improving VMC that are reported in literature. Finally, the chapter outlines the concept of using shared virtual reality applications as a communications medium, presents examples of CVEs, and reviews related work from that area.

Chapter 3 introduces the video-CVE prototype "cAR/PE!" (Regenbrecht et al. 2004), which combines VMC and CVEs, and which serves as the research platform for this work. Then, it explains the research approach pursued in this thesis to increase our understanding of human factors of video-CVEs. Finally, it gives an overview of how the research approach is realised in the four user experiments that are described in the four subsequent chapters.

Chapter 4 presents findings from the experiment "Desert Survival Game". There are no well established metrics that apply for comparing variants of VMC interfaces. Thus, the approach taken in this experiment was to trial subjective measures that have previously been used in similar cross-media studies for their sensitivity to discriminate between standard video-conferencing, video-CVEs, and face-to-face conditions. The considered measures included two social presence scales, and a questionnaire addressing subjective communication quality. Results were also explored for first indications, where video-CVEs push the collaborative experience further towards the face-to-face gold standard.

Chapter 5 presents findings from the experiment "Dream House". This experiment re-applied one of the social presence scales that was found to be useful in experiment "Desert Survival Game" to strengthen the previous result. In addition, it trialled a physical presence scale for its ability to discriminate between a standard video-conferencing system and a video-CVE, and investigated whether or not participants' physical presence ratings and social presence were related.

Chapter 6 presents findings from the experiment "Dogs & Owners". This study concentrated on the impact of the spatiality aspect of video-CVEs. In particular, it tested the hypothesis that face-to-face-like collaboration can be approximated by creating a sense of spatiality in VMC. This hypothesis was tested based on subjective questionnaire data and video analysis.

Chapter 7 presents findings from the experiment "Celebrity Quotes". This study concentrated on the impact of the transportation aspect of video-

CVEs. In particular, it tested the hypothesis that face-to-face-like collaboration can be approximated by increasing the level of transportation of a video-CVE. This hypothesis was tested based on subjective questionnaire data, video analysis, and a conversation transcript analysis.

Chapter 8 discusses the overall findings and the implications that can be drawn from them and proposes directions for future work.

Chapter 9 reviews the main findings of this work.

Chapter 2

Background

The research field of Computer Supported Cooperative Work (CSCW) investigates how collaborative activities and their coordination can be supported by means of computer systems. The aim of this thesis is to contribute knowledge to the CSCW area of real-time distributed groupware, that is, computer systems that allow multiple people to collaborate with each other at the same time but from different places. Understanding the challenges of real-time distributed groupware presupposes a fundamental understanding of interpersonal communication. This chapter establishes such an understanding and thus lays the foundations for this work.

Section 2.1 explains fundamental communication terminology and mechanisms involved in interpersonal communication and collaboration based on Shannon and Weaver's (1949) transmission model of communication. Section 2.2 highlights how different telecommunication media affect collaborative processes and illustrates the challenges designers of these systems are facing to provide people with the support they need to accomplish their remote collaborative goals. Section 2.3 focuses on video-mediated communication. It contains a review of a series of empirical studies that established a body of knowledge with regards to the value of video-conferencing as a medium to support remote collaboration. It also summarises new interface approaches and directions researchers are investigating in an attempt to improve VMC. Finally, Section 2.4 outlines the concept of Collaborative Virtual Environments (CVE) as a novel telecommunication medium.

2.1 Getting a grip on communication

This work involves human-to-human communication, so this chapter begins by explaining the human communication process itself. The three models presented are all based on Shannon and Weaver's (1949) transmission model of communication. They are contrasted with each other in the following sections. Section 2.1.1 starts with the original transmission model, which considers communication merely as an information processing problem. Section 2.1.2 presents Warren Weaver's adaptations to the original transmission model to also approximate interpersonal communication processes. Finally, Section 2.1.3 introduces a simple transmission model for interpersonal collaboration that further extends the previously explained transmission models.

Explaining the individual communication models with an explicit focus on the differences between them clarifies the terminology involved and establishes a good theoretical understanding of the matter.

2.1.1 Shannon and Weaver's model of communication

Human communication is complex and the research field that studies it is scattered. Scholars in the field of communication hold widely divergent views of what communication is and on the best way to study it. To date, there is no uniform theory that can account for all aspects of communication. Instead, Griffin (2005, page 21) argues that "it is reasonable to talk about a *field of communication theory*, in which a variety of different communication theories, models, and traditions coexist, all trying to answer questions of practical relevance that emerge when humans interact."

Shannon and Weaver (1949) developed one of the earliest and most influential communication models which considers the transmission of information from a source to a destination as the basic mechanism of communication. It is therefore referred to as the *transmission model* of communication (Figure 2.1).

According to Shannon and Weaver, a communication system comprises the following fundamental elements:

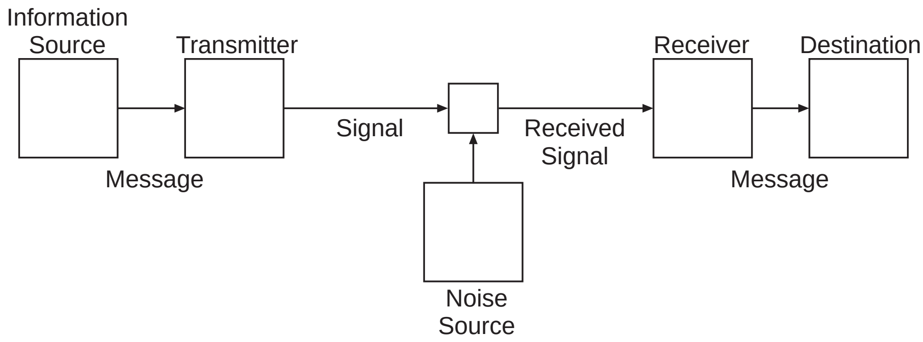


Figure 2.1: Shannon and Weaver's (1949) transmission model of communication.

The information source: Typically, the source is a person with a reason for engaging in communication.

The message: At the beginning of a communication process, the source formulates a message as a symbolic representation of an idea it wants to communicate to the destination.

The transmitter: The transmitter accepts a message and transforms it into a signal that can be sent to a remote receiver.

The channel: Once a signal is encoded, it is then sent through the channel.

The receiver: Signals sent via the channel are picked up by the receiver which converts them back into the message that can be perceived and understood by the destination.

The destination: The destination can now perceive and interpret the reconstructed message as presented by the receiver.

Noise: From the moment a message is formulated by the source until it is reconstructed at the destination, interfering noise causes loss of information. As a consequence, the message at the destination is always a distorted version of the message sent.

As an example, consider person A calling person B using a phone system: Person A is the *source*, person B is the *destination*. Person A formulates a *message*, consisting of a series of spoken words that have associated meanings. The microphone in the telephone acts as the *transmitter* and converts these spoken messages into electrical signals. These signals are then sent via the telephone *channel* (a set of wires) that is capable of carrying electrical impulses. The signals are then *received* by the telephone speaker at the destination, which transforms them back into sound waves that can be perceived and understood by Person B.

Communication fails if the received message does not carry the minimum information necessary to reduce uncertainty at the destination. In Shannon's view, information refers solely to the reduction of uncertainty. The amount of information a message carries is determined by the number of alternatives when selecting the message. The more open-ended and unpredictable the message, the higher its level of information. In this context information should not be confused with meaning. The relation between noise and information can be described by the simple equation:

$$Channel\ Capacity = Information + Noise$$

Shannon initially developed this model from an engineering perspective with the goal of maximizing the level of information a given channel could carry by minimizing the interfering noise. In his view, communication is simply an information processing problem. The impact a message can have on the destination, or whether the symbols that are transmitted from sender to receiver convey the desired meaning, was of no interest to Shannon. He writes, "the semantic aspects of communication are irrelevant to the engineering aspects" (Shannon and Weaver 1949, page 99).

With its technical focus, Shannon's model in its original form was neither capable of providing a general analogy of human communication, nor was it intended to do so. However, Warren Weaver, the co-author of "The Mathematical Theory of Communication", was more interested in the philosophical implications of the model and discussed those in an interpretative essay that follows Shannon's mathematical derivations in the same book. In his essay, he discussed the relevance of Shannon's communication model not only for the transmission of *symbols*, but also for the transmission of the desired *meaning* conveyed by the symbols and thus applied the concept of information loss to interpersonal communication. It might be due to this essay that Shannon's diagram of information flow appears in almost every communication textbook.

The next section explains an adapted transmission model, and highlights how the changes undertaken correspond to the general characteristics of interpersonal communication.

2.1.2 Interpersonal communication

Griffin (2005, page 52) defines interpersonal communication as the "process of creating unique shared meaning". There are a number of elements that set interpersonal communication apart from other forms of communication (e.g. mass communication). In interpersonal communication there are few participants involved, the participants are in close physical proximity to each other, there are many sensory channels used, and the feedback is immediate (Gouran et al. 1994).

With these definitions in mind, two main aspects emerge where Shannon's communication model does not fit:

First, interpersonal communication is dynamic and bilateral, and can therefore not be conceived as a linear and literal transmission of information from a sender to a receiver. Second, interpersonal communication is primarily about the exchange of meaning. A model therefore must also account for the semantic levels of communication. In this regard, Roszak (1986) criticises the unique way in which Shannon defines information as merely a measure of the difficulty in transmitting the sequences produced by some information source. Consequently, as Roszak points out, Shannon's model is incapable in distinguishing messages with valuable meaning from pure nonsense.

Figure 2.2 shows an adapted version of the Shannon-Weaver model which incorporates several modifications to remedy these shortcomings. A sender encodes an idea into a message and sends it via a channel to a receiver who decodes the message and tries to reconstruct an image of the idea. Communication is successful if the image created at the receiver side corresponds to the initial idea of the sender.

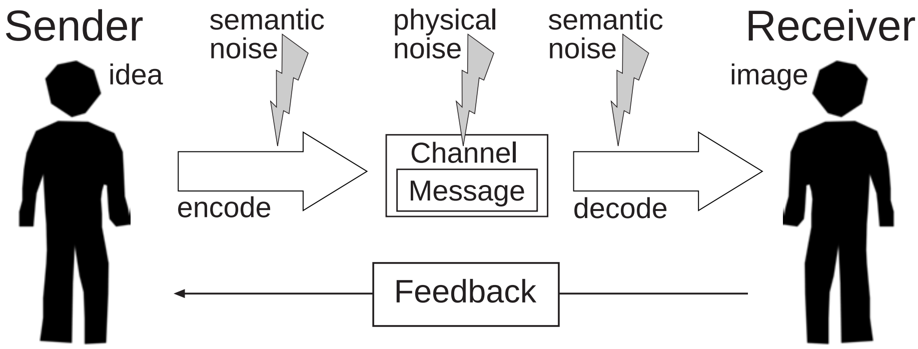


Figure 2.2: The Shannon-Weaver model of interpersonal communication, simplified from DeVito (1998, page 12).

Feedback: Knowing the success or failure of a sent message is important for the sender as it provides guidance when selecting the content and the encoding strategy of following messages. In face-to-face communication, the receiver can indicate successful decoding of a message into something meaningful by the occasional "mm-hm", "yes, I see", subtle head nods, smiles, and so on. Knowing that a receiver has been able to decode messages allows the sender to continue encoding in the same manner, or even to speed up. On the contrary, if a receiver indicates a potential decoding error with a surprised look or a "sorry?", "hold on", and so on, it gives the sender the chance to repeat, re-encode, and clarify. This self regulating mechanism between sender and receiver is accounted for in the adapted communication model by the addition of a feedback loop[1]. The inclusion of feedback significantly changes the conception of communication from being a static and one-way information transmission of information to being a dynamic system with the purpose of meaning transfer in which sender and receiver are mutually dependent.

[1] The concept of feedback was introduced by Norbert Wiener (1948) roughly at the same time in his work on cybernetics.

Including meaning: While the terms *encoding* and *decoding* of messages were merely used in the context of signal (symbol) conversion in the original model, the same terms are used in the adapted model within a higher semantic context. The sender is the encoder who converts ideas into messages. The receiver acts as the decoder who tries to reconstruct an adequate image of the initial idea by means of interpretation. In interpersonal communication, decoding and encoding are therefore referred to as complex cognitive (instead of technical) processes. Therefore, in a telephone call it is not enough for a receiver to only *hear* a message properly: he or she must *understand* it in order for communication to be successful.

Semantic noise: Including the semantic level into a model of communication introduces a new problem. What, for example, if the sender and the receiver speak entirely different languages? In that case verbal communication is bound to fail as the receiver is not capable of decoding the messages the sender encoded. This problem is accounted for in the adapted model through the inclusion of semantic noise at the decoding as well as the encoding side. At a basic level, in order for communication to be successful and to reduce semantic noise to a minimum, it is necessary that the sender and receiver share a common set of vocabulary, or more generally, a common set of meaning-to-symbol-associations. The decoding of messages can be tricky if they contain a certain level of ambiguity. In these cases, the decoding process involves a lot of interpretation where contextual factors such as the current situation, the relationship between sender and receiver, and personal experiences play crucial roles.

The Lasswell Formula: The impact of the relation between sender and receiver, as well as the impact of the choice of the appropriate channel, is of pivotal importance for communication processes. Building on the adapted Shannon-Weaver model, the sociologist Lasswell therefore recommended the consideration of five elements when studying the social aspects of communication (see Figure 2.3).

Lasswell's primary interest was in mass communication. Therefore he explicitly included the effects of communication as a separate element. How-

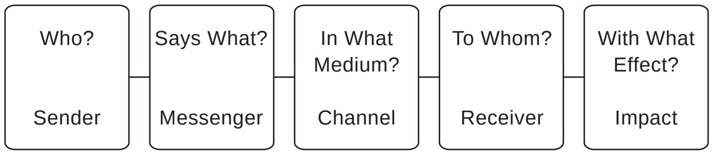


Figure 2.3: The Lasswell formula (Lasswell 1948).

ever, breaking down communication into the given five elements is also helpful for studying interpersonal communication. Guided by this framework, the following sections briefly outline selective findings from communication studies that are relevant to this thesis.

Message and channel: Messages in interpersonal communication can be composed as verbal or nonverbal expressions, which are then conveyed via the auditory and visual channels[2]. While verbal messages (e.g. spoken or written words) carry the formal content of a message, non-verbal messages add to and complete their meaning, provide feedback from listeners, and facilitate coordination. Argyle (1992, pages 8-12) lists the tone of voice, pauses, gestures, and gaze as non-verbal signals a speaker typically emits while speaking. Likewise, he finds vocalisations, gestures, facial expressions, posture, and gaze to be typical non-verbal signals mainly for the purpose of feedback on the side of the listener.

Eye-contact and gaze direction play a significant role in human social behaviour in general and in the coordination process during conversations in particular (Kendon 1967). As Argyle and Cook (1976) point out, speakers and auditors use gaze during face-to-face conversations to exchange and maintain roles, to regulate turn-taking behaviour, to signal attention or boredom, and to give and seek feedback in the form of short glances.

When engaged in a conversation, people are not always aware of the many non-verbal messages they are permanently encoding, sending, receiving, and decoding. Yet, according to Argyle and Dean (1965), the choice and the frequency of certain non-verbal messages are neither random nor independent of

[2] The olfactory and tactile channels also play roles in interpersonal communication. However, they are not relevant to this thesis and are therefore neglected here.

each other. Argyle and Dean (1965) suggest a theory in which an equilibrium for a certain level of *intimacy* is developed between people, where intimacy is a joint function of eye-contact, physical proximity, intimacy of topic, or smiling. There is a certain level of intimacy participants feel comfortable with and which they try to maintain. Therefore, if the intimacy equilibrium is disturbed because of the increase or decrease of one component, other components compensate to re-establish the equilibrium. For example, if two people increase the physical distance between themselves, they decrease their level of intimacy, which they can then re-establish by an increase in smiling or eye-contact (for example).

Sender and receiver: Communication styles differ between people. The impact of culture, power relations, social class, or gender have been studied in depth. Findings in that field reach far beyond the scope of this thesis. The next paragraphs can therefore only selectively show two types of differences that can be expected when comparing communication first between friends and strangers, and second, between men and women. Differences based on the familiarity and gender of communicators were selected based on their relevance during the remainder of this thesis.

As previously mentioned, for successful decoding of messages, sender and receiver have to share a common set of vocabulary or *common ground*, which is some shared information that can be taken for granted and that each assumes the other possesses (Argyle 1992, page 9). Between people who know one another well or who belong to the same group or community, or who work together, there is extensive common ground (Clark 1985). It can therefore be expected that the better people know each other, the easier it is for them to communicate. In contrast, when strangers meet for the first time, they have to get to know each other and build up a common ground during the course of the conversation. Communication patterns are then likely to be dominated by social protocols that reflect appropriate behaviour for politeness, respect, and friendliness. Sharing an idea with a stranger thus involves more work and is therefore less efficient.

Men and women tend to communicate in different ways, especially when talking to a same gender friend. There is a general consensus in the literature that women engage more in self-disclosure, and talk about more intimate topics to other women, while men avoid externalising emotion and instead are more oriented toward some external task, competition, or activity sharing. Wright (1982) summarises that "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 personalised knowledge of and concern for one another." Hall (1990) reviewed findings of several studies on non-verbal gender differences which seem to be aligned with this characterization. Hall's conclusions include that female faces are more expressive, that women smile more, and that women seem to gaze more at others than males do.

Effect: Lasswell introduced the question for possible effects or outcomes to the study of communication. An outcome of particular interest for the study of mass media is, for example, the degree to which a sender is able to persuade or deceive his audience.

In the context of cooperative communication, however, it is assumed that both sender and receiver have a common problem that they are trying to solve. The desired outcome of communication in this case would be that a message is understood and ultimately helps the sender and receiver to reach their common goal.

The next section explains the concept of collaboration as a special case of interpersonal communication.

2.1.3 Collaboration

Section 2.1.1 introduced the general Shannon-Weaver model of communication, Section 2.1.2 showed how this model can be applied to describe the process of interpersonal communication. This section focuses on collaborative interpersonal communication, the type of human-to-human interaction that is at the centre of CSCW research.

Collaboration generally refers to the process of people being engaged in joint, interdependent activities in order to achieve a common goal[3]. A classical example for collocated, synchronous collaboration would be a round-table meeting where participants discuss various documents and have to achieve a defined goal (find the best solution/decision for a problem) within a limited time frame. Collaboration, like interpersonal communication, is the process of creating unique shared meaning, where the body of accumulated shared meaning serves as the basis for joint action that leads to solving the task at hand. Figure 2.4 shows a simple model for collaboration, which is a further extension of Shannon-Weaver model of interpersonal communication discussed earlier.

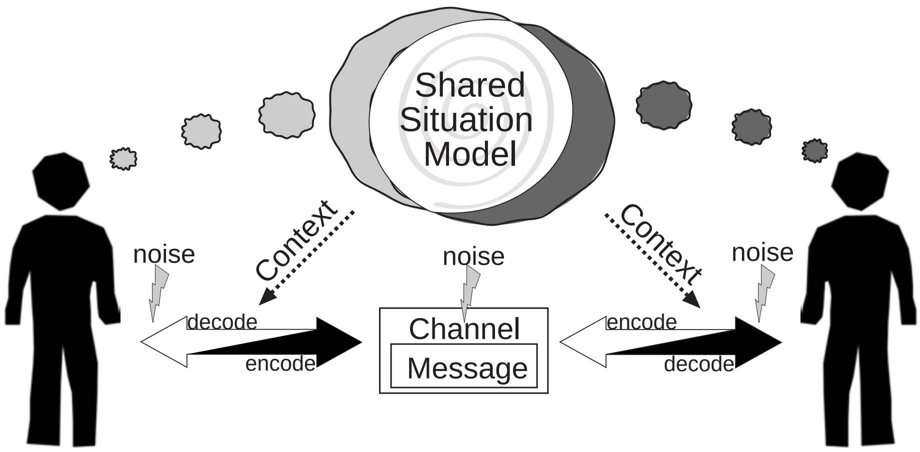


Figure 2.4: A simple model for collaboration.

Collaboration is driven by the exchange of messages. Participants engage in both sender and receiver roles and therefore both encode/send messages and receive/decode messages from each other, including feedback. During the process of collaboration, every participant develops his or her own mental *Situation Model* which contains the available information, possible alternative solutions, evaluation of alternatives, and so on. The largest part of the collaborative effort for participants is to expand each other's situation models and maximise overlap. (The overlap represents participants' shared understanding which is the common ground from where a solution for the

[3] For an in-depth review of different definitions of the term "collaboration", please refer to Hornecker (2004, Chapter 4).

task at hand can evolve.)

Grounding: Accumulating common ground is an interactive process and is fundamental to all collective actions (Clark and Brennan 1991). To succeed in grounding a message, that is to make it part of their common ground, it is crucial for participants to ensure that a message has been understood as it was intended. The way in which participants ground their messages depends on the individual situation.

- Verbal grounding: grounding is evident in spoken conversations. As previously mentioned, listeners may signal their understanding in form of positive feedback, also referred to as acknowledgements or backchannel responses. For other, more subtle forms of verbal grounding refer to Clark and Brennan (1991).

- Non-verbal grounding: Fussell et al. (2000) collected different types of visual (non-verbal) information and highlighted their relevance for three grounding subtasks. Table 2.1 is an adapted version of their table. It shows the different forms of body language people can use to coordinate each other's attention and give feedback. Fussell et al. (2000) also considered the often neglected role of shared objects and a shared work context on facilitating grounding.

Least collaborative effort: When trying to ground their utterances, participants strive to encode their messages in the most efficient and appropriate way possible. An important general rule that underlies the grounding processes is the *principle of the least collaborative effort*, according to which participants in conversations always try to minimise their collaborative effort – the work that both do from the initiation of each contribution to its mutual acceptance (Clark and Brennan 1991).

For example, being able to point at a visually shared object during a collaboration allows for a very easy, brief, and concise way of referencing through a deictic utterance (e.g. "this one") and therefore helps to reduce collaborative effort. It can therefore be expected that participants use indicative

	Type of Visual Information			
Grounding Subtasks	Participants' heads and faces	Participants' bodies and actions	Shared task objects	Shared work context
Establish joint focus of attention	Eye gaze and head position can be used to signal area of attention	Body positions and activities can be used to signal area of attention	Constrain possible foci of attention	Constrain possible foci of attention; disambiguate off-task attention
Monitor comprehension	Facial expressions and gestures can be used to give feedback	Approprietness of actions can be used to infer comprehension, or clarify misunderstandings	Change in state of objects can be used to infer comprehension, clarify misunderstandings	
Conversational efficiency		Gestures can be used to point to task objects	Visually shared task objects can be referred to with deixis and pronouns	Environment can help constrain domain of conversation

Table 2.1: Benefits of four types of visual information for three grounding subtasks (adapted from Fussell et al. (2000)).

gestures (pointing, looking, touching) whenever possible to refer to nearby objects they attend to, as gestures come at a lower collaborative cost than having to describe the objects verbally. However, Clark and Wilkes-Gibbs (1986) point out that in real conversations minimizing effort does not necessarily always lead to the formulation of the shortest and most appropriate utterances. Rather, starting with an improper utterance and then repairing it together with the other participant, or splitting up one complicated utterance into multiple smaller parts can involve less work than having to plan and encode one flawless message in the first place.

2.1.4 Criticisms of transmission models

Message-centred transmission models are attractive to several academic disciplines because of their simplicity, generality, and quantifiability. However, many communication theorists consider them misleading and misrepresentations of the nature of some aspects of human communication.

Reddy (1979), for example, argues the *conduit metaphor* of sending messages that *contain* units of retractable meaning is oversimplified as it does not reflect the complex cognitive processes that are necessary for sender and receiver to actually make sense of a message. A communication theory should conceive of meaning as being actively constructed rather than being passively extracted, and should therefore account for the powerful human ability of interpretation. Furthermore, Carey (1989) points out that communication models where the receiver is a passive target who merely has to unpack and accept mirrored images of ideas that exist in the sender's mind would allow the sender not only to share ideas but also to control and manipulate people. This, however, neglects that the receiver can interpret and evaluate the content of a message, and then decide whether to accept, ignore, or oppose it. These possibilities are not considered in transmission models. Other points of criticism include that transmission models generally overlook the context and the dynamic changes of the context in which communication takes place, and that they do no justice to the impact of the relationship between sender and receiver.

These issues must be kept in mind when applying the previously introduced transmission models in the remainder of this work.

2.1.5 Summary

The last sections explained communication in terms of the transmission of messages between a source and a destination. Starting with Shannon and Weaver's (1949) general model of communication, basic communication terminology was introduced. Successive sections then applied the principle of message exchange first to interpersonal communication, and second to collaboration between people.

The goal of interpersonal communication is to establish a shared understanding between sender and receiver. Communication between people is effective if shared, unique meaning can be created. Communication between people is furthermore efficient, if the creation of shared meaning involves the least possible effort. Communication is more efficient, if people can choose the easiest from different available communication channels to formulate their messages, and if they can easily monitor if the messages they sent were understood by others the way they were intended to.

Although conceiving human communication mainly in terms of exchanging messages is simplified and does not do many aspects of human communication justice, it serves as a useful model for the following sections, which will discuss how the exchange of messages is influenced by the use of different telecommunication media.

2.2 Mediated communication

If it is impossible for people to be at the same place at the same time, they can use telecommunication technology to bridge the distance that lies between them. A large variety of communication-mediating technologies are available. Section 2.2.1 characterises different types of media in the context of social presence and media richness theory, as well as in terms of their ability to support grounding mechanisms. Section 2.2.2 outlines theoretical considerations found in the literature that draw connections between media characteristics and media use. Section 2.2.3 then focuses on strategies that were implemented to mediate situation awareness in situations where distant persons collaborate in shared applications and online workspaces.

2.2.1 Media characteristics

Telecommunication media differ in their characteristics and have been categorised in numerous ways. Media can for example be coarsely distinguished as being synchronous (supporting communication in real time), or asynchronous (supporting non-real time communication) (Johansen 1988). More finely grained classifications characterise different media according to:

- their richness in terms of their ability to convey ambiguous information.
- the techniques they allow for grounding.
- their degree of salience of the interpersonal relationship in the interaction (social presence as a subjective quality of a communication medium).
- the degree to which they support a sense of "being together with another" (social presence as psychological state).

In the following, these four classifications and theories are explained in more detail.

Media richness

Daft and Lengel (1984, 1986, 1987) developed the concept of media richness in an attempt to match the appropriate type of medium with the level of ambiguity of a given task. Communication media have an intrinsic “richness” based on their support for instant feedback, the number of available channels and cues, language variety, and personalization. Daft and Lengel suggest that the richer a medium is, the better it can handle ambiguous situations which often include disagreements, confusion, or a lack of understanding.

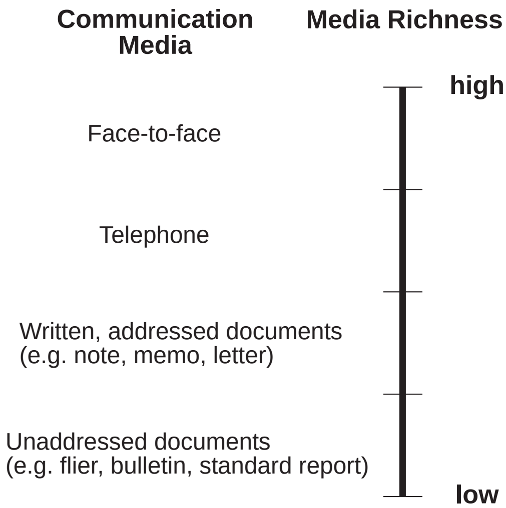


Figure 2.5: Richness of different communication media, adapted from Daft et al. (1987, page 358).

Figure 2.5 shows the hierarchy of a selection of communication media (Daft et al. 1987, page 358). Face-to-face is the richest communication medium. The telephone medium is less rich, because visual cues and body language are filtered out, putting emphasis on language content and audio cues to reach understanding. Telephone conversations, however, allow for fast feedback and are personal, because they allow for natural language. Asynchronous, written media are low in richness, because they do not provide rapid mutual feedback and inhibit any audio cues. Visual cues are reduced to those on the paper. Further differences in richness among written media are determined based on the level of personalization they support (addressed versus unaddressed documents).

Media and grounding

As noted in Section 2.1.3, achieving common ground is pivotal to all cooperative activities. When people interact with each other at a distance, the communication medium used has a considerable impact on the possible ways common ground can be established. Clark and Brennan (1991) therefore suggested that any medium that supports cooperative work can be characterised in terms of the ways it supports grounding. They provided the following framework containing eight constraints that a medium may impose on communication between two or more participants:

- Copresence: Participants share the same physical environment
- Visibility: Participants can see each other
- Audibility: Participants can use speech to communicate
- Cotemporality: Messages are received immediately after they are sent
- Simultaneity: Participants can send and receive messages at the same time
- Sequentiality: Participants' turns cannot get out of sequence
- Reviewability: Messages can be reviewed
- Revisability: Messages can be revised

By placing different media in this framework, differences as well as similarities between them emerge that help us understand and distinguish the ways in which different media may allow the acquisition of common ground.

Table 2.2 graphically underlines the different nature of synchronous and asynchronous media. It also shows where asynchronous media can potentially outperform synchronous media, namely in situations where it is beneficial if messages can be reviewed and revised ex post.

Medium	Copresence	Visibility	Audibility	Cotemporality	Simultaneity	Sequentiality	Reviewability	Revisability
Face-to-Face	•	•	•	•	•	•		
Videoconference		•	•	•	•	•		
Telephone			•	•	•	•		
Instant Messaging				•	•	•	•	•
Answering Machine			•				•	
E-mail							•	•
Letter							•	•

Table 2.2: Constraints of different media for grounding processes (adapted from Olson and Olson (2000)).

Social presence as a subjective quality of medium

In an attempt to determine the relative effectiveness of different media channels for social communication, Short et al. (1976) developed the "Social Presence Theory". It is based on Douglas's (1957) premise that every interaction between two persons involves an "interparty" as well as an "interpersonal" aspect, where the interparty component is task-focused, and the interpersonal component is concerned with developing and maintaining some personal relationship.

Morley and Stephenson (1969) subsequently argued that there is a balance between the interparty and the interpersonal component which is affected by a certain communication medium, or, more accurately, by the amount and the combination of non-verbal cues that a certain communication medium supports. In telephone conversations, for example, with the absence of the visual channel which is mostly associated with the interpersonal channel, this balance is shifted more towards the interparty side. Morley and Stephenson explain that, therefore, people talking over the phone are likely to be less concerned with the presentation of the self, and more task-oriented.

Short et al. believed that "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships is an important hypothetical construct... (page 65)", which they term "Social Presence".

Social presence is conceived to be a quantifiable property of a communication medium as perceived by a person who is using the medium. The absence or existence of certain verbal or non-verbal cues in a communication medium contribute to its social presence as they affect the interparty/interpersonal balance in one way or the other. However, as social presence is understood as a perceptual or attitudinal dimension of the user, it can not simply be "calculated" by adding the number of existing communication channels, but rather results from a user's "cognitive synthesis" thereof.

The semantic differential technique: Social presence, although a property of the communication itself, is subjectively experienced by a user in the form of a "mental set" towards the medium, and can therefore only be measured subjectively. For assessing social presence, Short et al. applied the semantic differential technique which was developed by Osgood et al. (1957). This technique was an attempt to measure the subjective meaning of any concept (real objects as well as abstract constructs) in the form of a discrete point within a semantic space that was spanned by dimensions empirically determined through factor analysis. In an attempt to find factors that determine the meaning of a particular concept, Osgood et al. applied multiple seven-step scales having a bipolar (verbal opposites) form and defined by adjectives. The underlying assumption is that thinking in terms of opposites is "natural" to the human species. Figure 2.6 shows an example.

good ____ ____ ____ ____ ____ ____ ____ bad

Figure 2.6: Example of a bipolar scale.

In the course of multiple studies that Osgood et al. conducted, they identified the dimensions *Evaluation* (e.g. good-bad), *Potency* (e.g. strong-weak) and *Activity* (e.g. active-passive) to be the three predominant dimensions that underlie human judgement and which span the human semantic space.

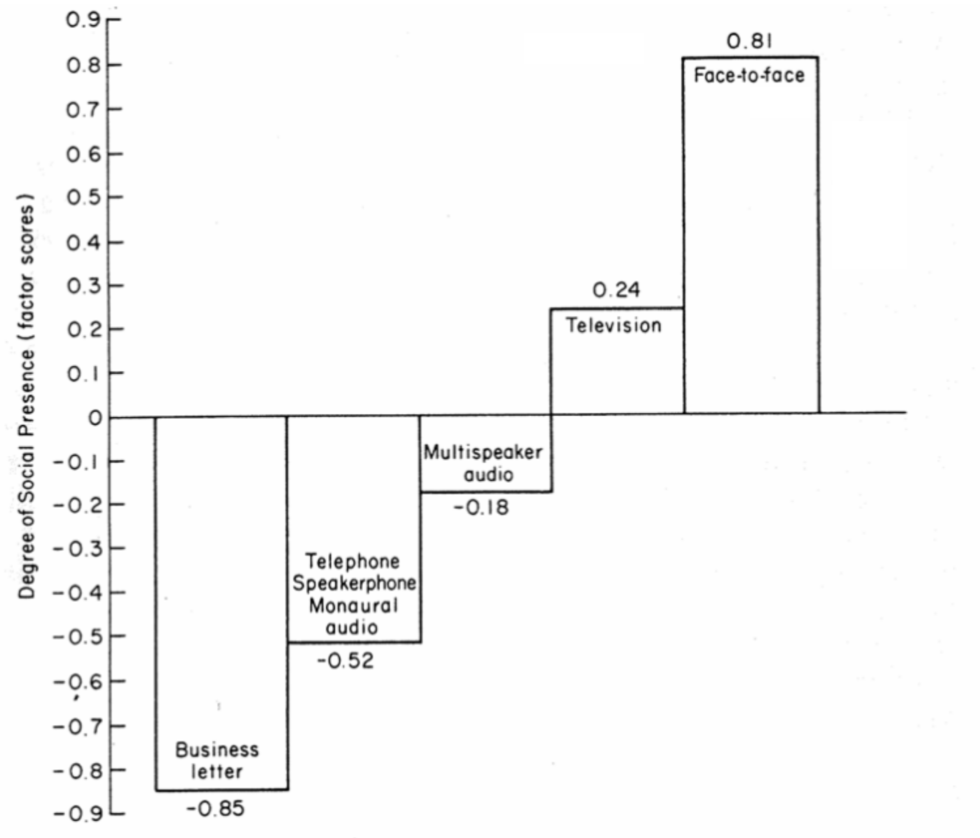


Figure 2.7: Social presence of several communication media, (Short et al. 1976, page 71).

Short and his colleagues also experimented with several semantic differential scales to assess some social and emotional capabilities of different telecommunication media. Among the numerous bipolar pairs applied, they found one recurring factor emerge from scales such as unsociable–sociable, insensitive–sensitive, cold–warm, and impersonal–personal which reflected the social presence of a medium. Media with a high degree of social presence were judged as being warm, personal, sensitive, and sociable.

The application of semantic differential scales to evaluate subjective media qualities is "the most commonly used measure of social presence"(Biocca et al. 2003, page 465). Figure 2.7 shows the ordering of social presence of several media as the result of two initial studies.

Face-to-face is rated highest in social presence, followed by visual, non-visual, and written media. Furthermore, the social presence ratings differ for “mono-aural speaker” and “multispeaker audio”, suggesting that social presence may vary even between two versions of the same medium. Short et al. (1976) therefore assume that social presence “depends upon not only the visual non-verbal cues transmitted, but also more subtle aspects such as the apparent distance or the realness of the other.”

Short et al. (1976) consider media with higher social presence not simply as the the *better* or *more effective* medium. Rather, the suitability of any medium depends on the fit between the social presence of a medium and the social presence required by the task. Therefore, the outcome of individuals communicating may be more affected by the social presence of the medium in tasks that involve a higher interpersonal or emotional component than in mere factual information transfer or simple problem-solving tasks.

Social presence as psychological state

Short et al. (1976) coined and popularised the use of the term “social presence” in the late 1970s. Since then, with the emergence of a presence research community (see Section 2.4.1) in the early 1990s, it has been broadened to include the “extent to which other beings (living or synthetic) also exist in the world and appear to react to you” (Heeter 1992), the “feeling that the people with whom one is collaborating are in the same room” (Mason 1994), and the “sense of being together” (de Greef and IJsselsteijn 2001) (for a detailed review of social presence definitions see Biocca et al. (2003)).

Biocca et al. (2001) recently attempted to integrate all existing definitions into one single, broader social presence theory. They argue that such a theory contributes to a wider understanding of social behaviour in mediated environments and allows researchers to better predict and measure differences among media interfaces.

Biocca et al. define mediated social presence as:

> “the moment-by-moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other (i.e., human, animate, or artificial being). Social presence varies from

a superficial to deep sense of co-presence, psychological involvement, and behavioural engagement with the other. As a global, moment-by-moment sense of the other, social presence is an outcome of cognitive stimulations (i.e., inferences) of the others cognitive, emotional, and behavioural dispositions." (Biocca et al. 2001, page 2).

Figure 2.8 depicts the three hierarchical dimensions of social presence which follow this concept and names their empirically determined factors.

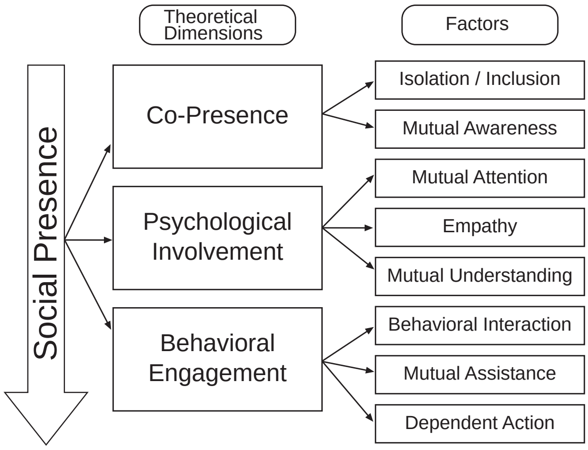


Figure 2.8: Factor structure of the Networked Minds definition of social presence (adapted from Biocca et al. (2001)).

In contrast to Short et al.'s (1976) conception of social presence as a subjective quality of a medium, this definition depicts social presence as a psychological, transient, and phenomenological state. Consequently, social presence is not assessed by judgement of the "medium" itself, but by judgement of the "experience" of another person through a medium. Therefore, social presence varies not only with the medium, but also with the knowledge of the other person, content of the communication, environment, and social context.

The Networked Minds (NWM) measure of social presence: Based on their definition of social presence, Biocca et al. (2001) also developed a questionnaire to measure it, the Networked Minds measure of social presence (see Table 2.3).

Factor	Items	Example
Isolation/ Inclusion	2	I often felt as if I was alone
Mutual Awareness	4	I hardly noticed another individual
Mutual Attention	8	I paid close attention to the other individual
Empathy	6	When I was happy, the other was happy
Mutual Understanding	6	The other understood what I meant
Behavioral Interaction	6	What I did affected what the other did
Mutual Assistance	4	My partner worked with me to complete the task
Dependent Action	2	The other could not act without me

Table 2.3: Example items of the NWM measure of social presence.

Biocca et al. (2001) validated the underlying theoretical factor structure of the NMW measure of social presence in a pilot study which compared the perceived social presence of the other in a standard video-conferencing (NetMeeting) condition and a face-to-face condition in a conversational task (Desert Survival Game). The results showed a significant difference between the video-conferencing condition and the face-to-face condition in six of the depicted eight sub-factors of social presence, but also revealed a problem with the internal consistency of the factors "Isolation/Inclusion" and "Dependent Action".

2.2.2 Media selection

People have a choice between many different communication media and choose and balance carefully and consciously when it is enough to send an email, when it might be better to call, or when it might be best to give someone a personal visit. Selecting an appropriate communication channel is primarily dependent on the type of task (Cockburn and Greenberg 1993).

Following up on the last section, this section summarises how the Media Richness Theory (MRT) and the principle of minimising collaborative effort explain why people choose certain media for certain tasks.

Media richness theory (MRT): Daft and Lengel's (1984) media richness theory proposed a natural fit between (1) information processing demands (the ambiguity of the situation) and (2) the information processing capabilities (the richness of the medium) in order to (3) optimise performance. In other words, to create shared meaning in the most efficient way, people choose rich media for equivocal tasks (e.g. if an inexperienced person has an IT problem and asks for help from an expert), and lean media for unequivocal tasks (e.g. if a new purchase order is sent to a supplier of a company). However, communication efficiency decreases, if the chosen medium is either too lean (ambiguity cannot be resolved) or too rich (unnecessary ambiguity is introduced). People intuitively anticipate the level of ambiguity of communications that lie ahead and choose a suitable medium accordingly.

To investigate if this assumption holds, researchers conducted numerous empirical studies into media selection, which produced contradictory results. While some studies seemed to be able to confirm the theoretically expected media preference (e.g. in Kraut et al. 1994), other studies produced weak or controversial empirical results (e.g. in Rice 1992, Lee 1994, Denis and Kinney 1998, Rice et al. 1998). Inconsistencies especially arose with regards to the use of newly emerged computer-based media like email or two-way chat. These media were not considered in the original media richness hierarchy and were added retroactively. According to critics however, the flexibility and adaptability of computer-based media inhibits their placement in any fixed hierarchy without taking into account the organizational context in which they are used. Other scholars furthermore tackled the assumption that media preferences exclusively depend on the ambiguity of a given situation, and suggested other interacting factors such as the features and usability of a given medium (El-Shinnawy and Markus 1998), or a complex set of social aspects in interpersonal relations (Harwood 2000).

Media choice and collaborative effort: According to Clark and Brennan's (1991) principle of least collaborative effort, people always try to ground with as little combined effort as needed. This principle also holds in mediated communication. Hence, for Clark and Brennan, participants' media choices are driven by the attempt to minimise the grounding costs associated with each medium relative to a given purpose. They list eleven types of grounding costs that contribute to the collaborative effort. Table 2.4 lists these costs together with "cheap" and "expensive" examples.

The different grounding costs interact with each other. For example, when someone uses an asynchronous medium like email, the time for message

Grounding Costs	cheap	expensive
Formulation	simple utterances	complicated utterances
Production	speaking and gesturing	typing and writing by hand
Reception	listening and looking	reading and waiting
Understanding	many contextual clues	missing contextual clues
Start-up	raising attention (f-t-f)	dialling a number (phone)
Delay	no delay	delay
Asynchrony	simultaneity	asynchronity
Speaker change	eye contact	verbally moderated
Display	showing, pointing, gazing	describing, talking
Fault	faults tolerated	faults not tolerated
Repair	self corrections	corrections through others

Table 2.4: Eleven grounding costs and examples (adapted from Clark and Brennan (1991).

formulation is not as restricted as it is in synchronous media. It therefore is easier for him or her to adequately phrase the messages and revise them until they meet his or her satisfaction. In turn, however, the receiver expects this and shows less tolerance if the email contains an error, which will then require even more time and effort to repair.

Summary: People's media preferences are dominated by the type of tasks they want to perform. While it is hard or even impossible for one theory to reliably predict fine grained media choices for all possible situations, it seems generally intuitively correct that rich media are preferred for tasks when the situation needs to be interpreted from one moment to the next, when information is incomplete and vague, where confusion and misunderstandings are likely to arise, and when the outcome of the situation remains unclear until the very end.

2.2.3 Remote collaboration in shared workspaces

Collaborators in face-to-face situations ground their ideas with the help of physical objects such as pen and paper, documents, photos, tables, or white-boards located in their common physical workspace (Bly 1988, Tang 1991, Fussell et al. 2000, Scott et al. 2003).

To provide similar visual support for collaboration at a distance, groupware systems often feature digital shared workspaces which graphically emulate tabletop or whiteboard surfaces. In real-time shared workspaces, participants manipulate or create task objects during the course of their collaboration, while all changes are immediately broadcasted to all other participants. Classic shared workspace applications include shared whiteboards for collaborative drawing activities, file viewers, or group editors (see Roseman and Greenberg (1996) and Prakash (1999) for examples). Empirical studies demonstrated the beneficial value of such tools. Whittaker et al. (1993), for example, found that the addition of a shared visual workspace to an audio channel lead to higher performance and an increased communication efficiency in a complex text editing task as well as a graphical task. The shared workspace also provided a permanent record of the task progress and provided a good a resource for monitoring and coordinating all ongoing collaborative actions.

Workspace awareness: Being able to coordinate one's individual activities relative to the overall collaborative process allows people to work together more efficiently. However, this presumes that participants know at all times what the others are doing. Dourish and Bellotti (1992) refer to such an understanding of the activities of others as "awareness". In the more specific scope of shared workspaces, Gutwin and Greenberg (1996) define "workspace awareness" as the up-to-the-moment understanding of another person's interaction with the shared space.

Gutwin (1997) presented a framework in which he identifies elements and categories of workspace awareness. The following table shows a selection of these elements that relate to maintaining workspace awareness on a moment-to-moment basis.

Maintaining workspace awareness in face-to-face situations usually comes at the cheap price of a short glance. However, keeping track of others' activities in distributed workspaces is challenging or sometimes even impossible, because groupware applications can only provide a fraction of the perceptual information that is available in a face-to-face workspace (Gutwin and Greenberg 1999). Individual participants are typically represented as tele-

Category	Element	Specific question
Who	Presence	Is anyone in the workspace?
	Identity	Who is participating?
		Who is that?
	Authorship	Who is doing that?
What	Action	What are they doing?
	Intention	What goal is that action part of?
	Artifact	What object are they working on?
Where	Location	Where are they working?
	Gaze	Where are they looking?
	View	Where can they see?
	Reach	Where can they reach?
When	Event History	When did that event happen?

Table 2.5: Elements of workspace awareness, adapted from Gutwin (1997).

pointers which do not convey any information about identity, gaze, view, or reach. Awareness support is considered a pivotal factor in remote collaboration and groupware usability. Therefore, researchers in the area of HCI and CSCW have explored various strategies to integrate awareness mechanisms into shared workspaces.

Supporting workspace awareness: One approach for increasing awareness is to restrict the possible actions others can perform. For example, participants can be assigned specific roles which describe an individual's relationship to the shared task objects and to other participants (Dourish and Bellotti 1992). Roles (e.g. author, reviewer, moderator) are typically linked to a predefined and limited set of operations which can be performed. Overall awareness then increases as the uncertainty about others' activities shrinks. Another example that increases awareness by restriction is the WYSIWIS (What You See Is What I See) multi-user interface abstraction. Shared workspaces that are designed in a strict WYSIWIS way always display the identical view of the workspace for all participants. That way every participant can be sure that all the others can see the identical view at all times. Restrictive concepts like this can be successfully applied in simple applications. However, they have been proved to be too limited and too inflexible for the demands of more complex collaborative tasks (Stefik et al. 1987, Gutwin and Greenberg 1998).

Another approach is to explicitly recreate relevant awareness information and provide this information in form of additional workspace interface features or "awareness widgets". For example, buddy lists can indicate the presence of other participants. Telepointers can be enhanced with name tags or icons to display the identity of the participants they represent. In a "relaxed WYSIWIS" workspace (one that allows participants to independently navigate through a shared workspace) the actual views of others can be illustrated by view rectangles ("teleports") (Beaudouin-Lafon and Karsenty 1992) or multi scrollbars (Roseman and Greenberg 1996). Activity levels of others can be indicated through the use of tangible ambient displays, such as the "activity balloon"(Prinz 1999).

Awareness mechanisms improve the overall usability of shared workspaces (Gutwin and Greenberg 1999). However, providing an adequate amount of awareness information that is not too restrictive or too obtrusive during remote interactions remains one of the main challenges for groupware designers.

2.3 Video-mediated communication

The first video-mediated communication (VMC) which involved transmitting a real time video of remote participants in addition to audio originated over fifty years ago. In those days, the vision for VMC was enthusiastic: the ability to see a remote person while talking to him or her would surely be something truly desirable and enriching for the daily lives of everybody. Pioneers were convinced that transmitting audio and video would have the capability to fully replace face-to-face situations, abolishing the burden of travel. With the introduction of the AT&T "Picturephone" in the 1970s, market forecasts predicted that soon all normal telephones would be replaced, and that, within a decade, the majority of all meetings would be mediated electronically (see Egido (1988) for a detailed historic review).

However, the picturephone failed dramatically, despite repeated marketing campaigns and in stark contrast to all forecasts. The marketed portrayal of video-conferencing as a direct replacement for face-to-face meetings proved to be exorbitant and utopian. Furthermore, it was obvious that video-mediated communication entailed a lot of subtle but crucial characteristics that were overlooked, ignored, or simply not well understood at that time.

Since then, the vision for VMC remains. However, in an attempt to find more efficient ways to use video as a communication medium, researchers have tried to identify existing problems and point at ways of overcoming these problems. Over the last thirty years, numerous scholars have studied people who communicate through video either in the lab or in the field and have built up a comprehensive body of knowledge.

Section 2.3.1 reviews a series of empirical studies which form the body of our knowledge about this medium to date. Then, Section 2.3.2 gives an overview of a number of new approaches to VMC that attempt to overcome some of the existing limitations.

2.3.1 Review of empirical work in video-mediated communication

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 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 below).

Table 2.6 lists the studies reviewed. The media conditions this review focuses on are face-to-face (FtF), audio-video (AV), and audio-only (AO). The conditions that are included in each listed study are indicated by a dot in the according field. Some of the mentioned studies comprised additional

	Conditions			
Task type	FtF	Video	Audio	Reference
Trust and Deception		•	•	Wichman (1970)
Negotiation	•	•	•	Short(1974)
Source - Seeker task	•	•	•	Chapanis (1975)
Brainstorming	•	•	•	Williams (1975)
Discussion	•		•	Rutter and Stephenson (1977)
Discussion	•	•	•	Rutter et al. (1981)
Real project		•	•	Tang (1992)
Discussion	•	•[a] •[b]		Sellen (1992)
Information Exchange	•	•[c] •[d]		O'Conaill et al. (1993)
Collaborative Design	•	•	•	Olson et al. (1995)
Informer - Follower task		•[e] •[f]	•	O'Malley et al. (1996)
Informer - Follower task		•[g] •[h]	•	Doherty-Sneddon et al. (1997)
Decision making		•	•	Daly-Jones et al. (1998)
Informer - Follower task		•	•	Veinott et al. (1999)
Discussion of photos		•	•	de Greef and IJsselsteijn (2001)
Informer - Follower task		•[i] •[j]	•	Monk and Gale (2002)
Trust and Deception	•	•	•	Bos et al. (2002)

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.

Table 2.6: Cross-media studies by task and conditions compared.

conditions which are neglected here to keep a uniform format. Some studies involved two different versions of the audio-video condition. This case is 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 the direct comparison. Gutwin and Greenberg (1999) distinguish between three types measures:

1. *Product measures* evaluate collaborative outcomes, considering both time and quality.

2. *Process measures* examine the efficiency of collaborative activities by analysing speech and interaction patterns of participants.

3. *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 Table 2.6 with regards to each of these measures.

Product measures

Product measures assume that the differences of communication media lead to measurable differences of the collaborative outcomes. The nature of these outcomes depend 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 (Chapanis 1975)). For these type of tasks, 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* (Reitman 1965, pages 148–156). Design tasks such like the one used in Olson et al. (1995) 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.

Another frequently used experimental task is the "Map Task" (O'Malley et al. 1996, Doherty-Sneddon et al. 1997, Veinott et al. 1999). Here, two participants get two slightly different versions of a terrain map. The map of one participant (informer) additionally shows a 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)" (Luce and Raiffa 1957) 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 receive a pay-off. Every player's ultimate goal is to maximise his/her pay-offs over the course of several rounds.

Table 2.7 shows the product measures applied and the results obtained in the studies considered in this review of cross-media studies.

Product Measures	**Findings**	**References**
Completion Time	FtF=AV=AO	Chapanis (1975)
	AV=AO	Daly-Jones et al. (1998)
Quality of Outcome	FtF=AV=AO	Williams (1975)
	FtF=AV>AO	Olson et al. (1995)
Reproduction Accuracy	$AV^e=AV^f=AO$	O'Malley et al. (1996)
	$AV^g=AV^h=AO$	Doherty-Sneddon et al. (1997)
	AV>AO *	Veinott et al. (1999)
	AV=AO **	Veinott et al. (1999)
	$AV^i=AV^j=AO$	Monk and Gale (2002)
Negotiation Results	AV≠AO	Wichman (1970)
	FtF=AV≠AO	Short (1974)
	FtF=AV=AO	Bos et al. (2002)

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.

Table 2.7: Cross-media studies: product measures and results.

As can be seen in Table 2.7, 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 Olson et al. (1995), who found a marginally worse quality in the design solutions that were created in the audio-only conditions.

The same pattern emerged for the results of the Map Task. The accuracy of the reproduced path did not differ across various audio-video and audio-only condition, with the noticeable exception being observed by Veinott et al. (1999). They 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 (1970) 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. (2002), who compared the outcomes of thirty rounds of a PD-like task between face-to-face, audio-video, as well as audio-only condition. 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 surprise the outcomes of audio-only and audio-video conditions were almost identical to each other over the whole course of the thirty rounds.

Short (1974) asked participants to negotiate a fictitious situation, where one participant represented a standpoint 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 participants, 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. (1996, page 126) 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 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.

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, the speaker turn frequency and length, turn taking behaviour, overlapping speech, interruptions, or the number of questions.

Table 2.8 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 experiment task caused a greater variance in results than that caused due to 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 and rather chaotic, whereas video-mediated and audio-only conversations are more formal and rigid. Spontaneous conversations are dynamic with frequent speaker changes, frequent interruptions, and overlapping speech. Formal conversations in contrast are characterised by fewer but longer "lecture-like" turns, hindered turn switching, fewer interruptions, and less overlapping speech. O'Conaill et al. (1993) found the formal character to be particularly articulated if the mediated audio suffers from a delay.

Olson et al. (1995) 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 partici-

Process Measures	**Findings**	**References**
Number of Words	AV[e]=AV[f]>AO	O'Malley et al. (1996)
	AV[g]>AV[h]=AO	Doherty-Sneddon et al. (1997)
	FtF=AO	Rutter and Stephenson (1977)
	AV[i]<AV[j]=AO	Monk and Gale (2002)
Number of Turns	FtF=AV[c]>AV[d]	O'Conaill et al. (1993)
	FtF=AV[a]=AV[b]	Sellen (1992)
	AV=AO	Daly-Jones et al. (1998)
	AV[e]=AV[f]>AO	O'Malley et al. (1996)
	AV[g]>AV[h]=AO	Doherty-Sneddon et al. (1997)
	AV=AO	Veinott et al. (1999)
	FtF=AO	Rutter and Stephenson (1977)
Turn Length	FtF=AV[c]<AV[d]	O'Conaill et al. (1993)
	FtF=AV[a]=AV[b]	Sellen (1992)
	AV=AO	Daly-Jones et al. (1998)
	FtF=AO	Rutter and Stephenson (1977)
Switching Times	FtF<AV[a]=AV[b]	Sellen (1992)
Handover by Name	FtF=AV[c]<AV[d]	O'Conaill et al. (1993)
Overlapping Speech	FtF=AV>AO	Rutter et al. (1981)
	FtF=AV[c]=AV[d]	O'Conaill et al. (1993)
	FtF>AV[a]=AV[b]	Sellen (1992)
	AV=AO	Daly-Jones et al. (1998)
	FtF>AO	Rutter and Stephenson (1977)
Interruptions	FtF=AV>AO	Rutter et al. (1981)
	FtF=AV[c]=AV[d]	O'Conaill et al. (1993)
	FtF>AV[a]=AV[b]	Sellen (1992)
	AV[e]=AV[f]>AO	O'Malley et al. (1996)
	AV[g]=AV[h]>AO	Doherty-Sneddon et al. (1997)
	FtF>AO	Rutter and Stephenson (1977)
Relative time spent …		Olson et al. (1995)
… 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;

Table 2.8: Cross-media studies: process measures and results.

pants used visual cues during their conversations which allowed their verbal conversations to be more task-focused and thus more efficient. Monk and Gale (2002) also demonstrated that the provision of full gaze awareness in video communication could reduce the number of words spoken to one half. Monk and Gale (2002) 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. (1997) and O'Malley et al. (1996) 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,...). 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.

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 somewhere in between audio-only and face-to-face communication. Yet, Sellen (1995) and Williams (1977) see the bigger resemblance between video-mediated conversations and audio-only communication rather than between video-mediated communication and face-to-face conversations.

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 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 2.9.

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. (1999), who found that a group of native English speakers perceived audio-only to be more efficient than audio and video, very much to the experimenters' own surprise.

Tang (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.

Satisfaction Measures	Findings	References
Preference	AV>AO	Daly-Jones et al. (1998)
	AV>AO	Tang (1992)
	AV[c,d]>AO	O'Conaill et al. (1993)
	FtF>AV>AO	Olson et al. (1995)
Subj. Comm. Efficiency	AV>AO *	Veinott et al. (1999)
	AV<AO **	Veinott et al. (1999)
Subjective Quality of Outcome	FtF>AV>AO	Olson et al. (1995)
Subjective Effort	AO>AV	Daly-Jones et al. (1998)
Interpersonal Awareness	AV>AO	Daly-Jones et al. (1998)
Ability to take control of the conv.	FtF>AV[a]=AV[b]	Sellen (1992)
Subjective Interactiviy	FtF>AV[a]=AV[b]	Sellen (1992)
Selective Attention	FtF>AV[a]>AV[b]	Sellen (1992)
Knowing when others were listening	FtF>AV[a]=AV[b]	Sellen (1992)
Social Presence	AV>AO	de Greef and IJsselsteijn (2001)

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; *Participants were non-native English speakers; **Participants were native English speakers.

Table 2.9: Cross-media studies: satisfaction measures and results.

Video facilitated the communication as gestures could be used. Also, while talking, users could see each others 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 to talk. 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. (1993) 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. (1995) furthermore perceived the quality of their outcome to be superior compared to the resulting outcomes in the audio-only condition. This is particularly interesting as this subjective opinion was not confirmed by the expert jury.

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

Summary of empirical work

Studies that directly compare audio-video communication with audio-only and face-to-face conversations are the most common attempt to discover existing differences which inform about potential benefits as well as shortcomings of being able to see the video of a remote person during remote encounters. However, evaluating video-mediated communication in that way 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. However, they are sensitive to the experiment 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:

1. *Video can add value to audio:* the degree to which video is of beneficial value in terms of better outcomes or communication efficiency is first and foremost determined by the type of collaborative task. Clearly, there are tasks where the addition of video is crucial and favourable.

2. *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. (1993) made clear that any fluidity and efficiency of communication processes crumble immediately with poorer audio quality. Any expected advantages through the addition of video relies 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 (Whittaker 1995).

3. *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.

2.3.2 Improving video-mediated communication

Besides demonstrating the potential benefit of video when added to audio, the reviewed studies also revealed substantial differences between remote collaboration using VMC and working together in a face-to-face situation, with the latter being the more preferred, the more effective, and the more efficient communication mode.

Research that seeks to improve video-mediated communication use three fundamentally different approaches:

1. Increasing the social presence of VMC.

2. Using video-as-data to convey critical information about the task rather than talking heads.

3. Using video to enable informal communication.

The next sections will highlight the problems that drive each of these approaches and demonstrate how researchers have attempted to overcome these shortcomings through the design of novel VMC-interfaces.

Increasing social presence in video-mediated communication

At the center of this approach is the question how missing non-verbal communication cues can be re-introduced in VMC. Researchers who follow this approach assume that the more non-verbal cues are supported, the more VMC will converge with face-to-face communication, which will ultimately create a sense of the remote person being actually present at a remote place. Missing non-verbal cues that have received the most attention in VMC research include the ability to establish eye-to-eye contact, gaze awareness, as well as proximity behaviour.

The following sections will give an overview of experimental VMC systems that demonstrate how one or more of these non-verbal cues can be conveyed over a distance.

Eye contact: The important role of eye contact in face-to-face communication is well understood (Argyle and Cook 1976, Kendon 1967). People look at each other to signal attention, just before they finish their utterance, and whenever they seek feedback from others. Because of an offset between the camera and the displayed video, the establishment of true eye-contact is impossible in standard videoconferencing setups (Gaver 1992). This may lead to problems in the coordination of turn taking between remote participants. Several interface approaches demonstrated how this shortcoming may be overcome.

- *Video tunnels:* a video tunnel setup uses a half-transparent mirror to reflect the video of a displaced monitor while allowing a camera that is mounted behind the mirror to capture video from virtually the same location as the video is displayed. The idea of video tunnels originated in the area of teleprompter broadcasting and was first applied for two-way video-conferencing in the late 1980s (Acker and Levitt 1987, Buxton and Moran 1990).

- *Hydra:* Sellen (1992) designed the video-teleconferencing system "Hydra" to improve awareness and selective attention in remote multi-party conversations. In Hydra, the video of each remote participant is shown on little terminals that integrate a small screen, a camera mounted closely to the screen, and a speaker. Each of these units are positioned spatially separate from each other on each participant's desk. The videos that are captured with each terminal are sent to the participant whose video is shown on the corresponding screen. Every participant can see the videos of the others as seen through the "eyes" of his or her terminal. Consequently, whenever two people look at each other's terminals they can establish quasi eye-contact, while the others see their faces from their terminals and therefore with a much bigger offset. User tests revealed that subjects had a better sense if others were listening or attending to them when they used "Hydra" compared to a standard video-conferencing system that showed the videos of all four participants together in one screen.

- *Multiview:* Nguyen and Canny (2005) presented a "spatially faithful" group-to-group videoconferencing setup which allows mutual eye contact between all members of two remote groups of three people. Like Hydra, MultiView provides participants with their individual view, captured by a camera at the corresponding position the remote end. The view separation for the three co-located persons is realised by using a projector per person which projects onto a screen of retro-reflective material, which reflects light always back to its source. As a result, every participant of the co-located group is able to see his or her own perspective of the remote scene, although all are facing the same screen. If cameras and projectors are carefully aligned to match the actual view of the participants, any two members of the remote groups can thus establish eye-to-eye contact.

Gaze awareness: Gaze awareness is knowing where someone else is looking and is useful because it allows remote participants to establish eye-contact. In addition, in the case where two remote participants are collaborating in a shared workspace, full gaze awareness refers to knowing what object the other person is looking at. Gaze awareness is an important non-verbal resource during conversations, as it allows participants to infer what others are attending to (Daly-Jones et al. 1998, Monk and Gale 2002). People are generally very good at estimating what someone else is looking at as long as the gazer and the gazed at object are visible (Gale and Monk 2000). However, in conventional video-conferencing systems the display of the shared workspace is spatially decoupled from the interpersonal space (video) of the remote participant, which makes gaze awareness impossible. As a consequence, the gaze direction of the remote participant as it appears in the video does not inform about what he or she is actually looking at. In an attempt to integrate shared workspaces with interpersonal spaces in a spatially consistent way that enables full gaze awareness, the following interface approaches have been suggested:

- *Clearboard:* Ishii at al. (1992, 1993) demonstrated a way to support gaze awareness over a distance by seamlessly combining shared workspace and the interpersonal space with the "ClearBoard" prototype. The main idea behind it goes back to Engelbart and English (1968), who had already experimented with superimposing a computer-generated display over the video signal from a remote camera. Clearboard utilises half-silvered mirrors with an integrated transparent digitizer sheet that accepts pen input for natural drawing while the drawing and the video of the remote person is back-projected onto the same screen. This is done in a spatially consistent way, which allows users to infer from the gaze of the other person in the video which objects on the drawing board he or she is looking at. That way, people can use ClearBoard in a natural and almost face-to-face way similar to "talking through and drawing on a big transparent glass board".

- *Gaze:* Vertegaal (1999) proposed a solution for supporting gaze awareness that required significantly lower setup costs. In their "GAZE" groupware system, so called "personas" which are 2D images or video of every participant are displayed as if they were arranged in space around a virtual table. Every participant can see the view of his or per persona. In addition, an eye tracker detects the fixation points of the participant on the screen. This information is then used to orient the virtual persona towards the source of attention, giving others the chance to infer where participants are currently looking. Vertegaal et al. (2003) extended the initial "GAZE" system later in "GAZE-2", which combined the concept of video-personas with more elaborate multi-camera and video-tunnel technology.

Proximity and interpersonal space: In face-to-face situations, proximity serves as an important non-verbal communication resource. People subconsciously perceive and differentiate between intimate, personal, and social space, and adapt their interactions accordingly (Hall 1969, pages 113-125). A close friend for example is tolerated in someone's personal space, whereas talking to a stranger usually occurs from a greater distance. Standard VMC does not support a sense of continuous interpersonal space between speakers and therefore inhibits a concept of negotiated mutual distance (Sellen 1992).

- *Hypermirror:* Morikawa and Maesako (1998) re-introduced some spatial relationships between remote participants in their system "HyperMirror". The HyperMirror essentially is one large display which integrates full-body videos of both local and remote participant into the same two-dimensional reference frame following the strict WYSIWIS paradigm. People looking at the HyperMirror see their own "mirror images" blended with the videos of the remote participants, who then appear to be close and within reach. Observations showed that users tried to behave in a way that resembled the proximity behaviour in co-located situations, such as avoiding invading the personal space of virtual others as it appeared on the HyperMirror.

- *Tele-Portals:* The vision for tele-portals is to merge distant spaces in a spatially consistent and seamless way, so that one continuous space, half local, half remote, half real, half virtual, emerges. In this way spatial relations between the participants are preserved which gives rise to additional non-verbal communication cues including proximity behavior, gaze awareness, and the ability to point at any object that is located in the shared space. One lasting research project that follows this vision and explores its technical feasibility is the "Office of the Future" (Raskar et al. 1998, Chen et al. 2000). Other experimental teleconferencing systems include "MAJIC" (Okada et al. 1994) or "TELEPORT" (Gibbs et al. 1999) as well as table-based systems such as the "VIRTUE" immersive 3D video-conferencing system (Kauff and Schreer 2002). Also advanced commercially available integrated video-conferencing solutions such as the system "Halo[4]" or the "Telesuite [5]" show similar design attempts to create the illusion of one continuous collaboration space.

- *Telepresence-Robot:* Jouppi (2002) demonstrated a further solution for the re-establishment of spatial properties between remote participants. He presented the prototype of a life-sized tele-operated robot which is supposed to visit remote locations embodying the person who is actually controlling it from a distance. A front-view as well as a side-view video of the remote person's face is displayed on the robot's "head", giving access to facial expressions as well as giving an idea about the actual view of the person at the other end. With its physical, life-size appearance, spatial and proximity behaviours can be adopted from normal face-to-face communication. The robot can for example be approached by others in order to attract "its" attention, or it can assume a position in a round table meeting or in a group that forms spontaneously around the coffee machine.

Using video as data

In standard VMC, video is typically used to show a head-and-shoulder portrait view of the remote person in an attempt to simulate a face-to-face situation. However, such a static "talking head" view is not always tailored to the conversational needs of people collaborating at a distance and therefore does not fully exploit the capacity of video to support remote collaboration.

---

[4] www.hp.com/halo/, last accessed June 2007.

[5] www.pangeair.com/, last accessed June 2007.

Video-as-data: Whittaker (1995) advocates the application of "video-as-data", that is, providing information about a shared workspace and its objects rather than the interactants. This is more likely to facilitate the establishment of common ground in remote collaboration and will therefore lead to a higher task performance than the display of talking heads only could afford. Nardi et al. (1993) investigated the use of video-as-data in neurosurgery, a domain where video has long been successfully deployed as data in a co-located setting. They found video-as-data serves a number of highly varied functions ranging from coordinating the fast exchange of surgical instruments to maintaining attention and focus and even educating medical personnel. From these observations, they also point out that although video-as-data does not facilitate direct interpersonal communication, the rich contextual information provided by the video often eliminates uncertainty and with it the need for engaging in any communication in the first place and suggest that these strengths of video-as-data for communication should not be overlooked in the design of efficient groupware systems.

Multiple camera views: Gaver (1992) disclosed another shortcoming of "talking head" videos. He pointed out that people in face-to-face situations constantly explore their environment by actively moving and looking around. A single static camera view as applied in conventional "talking head" video-conferencing systems, however, leads to a feeling of being "paralysed" at the remote location. In an attempt to allow more perceptual exploration at the remote end, Gaver et al. (1993) therefore presented the experimental system Multiple Target Video (MTV) which featured multiple camera views of the remote site among which the users could switch. The system provided five different views: a face-to-face view, a bird's eye view, a desk view, a doll house view, and an in-context view. Pairs of users were asked to work on two collaborative tasks[6] while experimenters recorded the usage of the different views. The results obtained for the given (visual) tasks showed that users chose the face-to-face view in only 11% of the total time, mainly for short glances to assess the other's engagement and mood. For most of the time,

[6] In the first task the participants were asked to draw a sketch of the partner's office; in the second task they were asked to select and arrange furniture in a doll house.

however, participants preferred to maintain views which showed objects that were related to the task. This result supports the claim that a static view of the other person's face may only deliver a fraction of the information which is directly relevant to collaborative processes and that task space video may be much more helpful.

Head mounted cameras: The concept of transmitting the objects and scene as seen by a remote person wearing a head mounted camera has been explored with regards to mobile settings such as disaster operations or remote repair tasks (Kraut et al. 1996, Fussell et al. 2000, Fussell et al. 2003). Although providing the actual view of a remote person was found to change the communication behaviour to some degree, with its current technological limitations the overall efficiency and performance could not be improved compared to audio-only or a bird's-eye overview video.

Using video to support informal communication

Organisations use VMC mainly to support remote collaboration in formal meetings. That means that use is typically planned and structured, with pre-scheduled start and end times. To initiate ad-hoc communication, a start-up request has to be accepted at the other end first. In both approaches, every minute of an established video connection counts. Attention and efforts are directed towards the primary task either to maximise the collaborative outcome of a given time frame or to minimise the time that it takes to reach a certain collaborative goal together with the remote party that was contacted.

However, research studies comparing long term collaboration between co-located and remote teams underline that one of the pivotal advantages of working in close proximity is the opportunity for informal social contact (Heath and Luff 1991, Olson and Olson 2000). In co-location, informal communication inevitably occurs during chance encounters in the hallway, before and after meetings, or during lunch and coffee breaks. These short and seemingly unimportant conversations serve an important role as they allow members of a team to stay aware of each other's work activities and social life and help to create a sense of community with its own established common ground.

A "Media space" is a computer-controlled network of audio-video equipment used to support collaboration (Gaver et al. 1992). Media spaces address the lack of informal communication in VMC by providing a permanent video link between remote sites. Several projects explored this concept.

- *Xerox PARC and EuroPARC:* The first Media Space evolved out of a self-experiment of a research group at Xerox PARC whose members were based in two sister laboratories that were located more than 1000 km apart from each other within the US (Bly et al. 1993). In an attempt to create an environment that would not only allow the remote members to work but also to "be" together, a permanent audio-video link was established between the two common rooms of the two locations. This helped maintain peripheral awareness of activities across both sites, and led to informal interactions like chatting and chance encounters (see Olson and Bly (1991) for a review). Following this example the next media space, "RAVE" (Gaver et al. 1992), was created at EuroPARC in Cambridge, England, in an attempt to collapse the separation between the colleagues at PARC and EuroPARC across the Atlantic Ocean. In addition, RAVE also connected people that worked in the same building, but were based in different rooms or on different floors. Each of the 30 staff offices as well as the common areas were equipped with cameras, speakers, and monitors which, once connected, continuously fed audio and video streams into the media space data network. The RAVE user interface was designed in a way that would allow interactions with different levels of engagement, ranging from unobtrusive one-way glances to see if a person was sitting at his or her desk to highly-focused two-way audio-video connections. Two co-workers could furthermore "share" their offices by establishing a permanent two-way video-link to provide a peripheral awareness about the other person's presence and activities.

  A live video connection between every member of EuroPARC and XeroxPARC was not feasible due to the high costs involved. Instead, to save bandwidth, still images which were updated only about every five minutes were distributed between the sites through the shared application Polyscope (Borning and Travers 1991) and later Portholes (Dourish and Bly 1992).

- *Cruiser and VideoWindow:* Researchers at Bellcore developed the system "Cruiser" to enable unplanned, informal, casual interaction (Root 1988, Cool et al. 1992, Fish et al. 1992). The design of Cruiser was centered around the basic concepts of social browsing, a virtual workplace, and interaction protocols, leading to a simulated experience of walking down a hallway, glancing into others' offices and taking advantage of chance encounters.

  "VideoWindow" (Fish et al. 1990) is another video-based system to support informal communication that was developed by Bellcore. The VideoWindow teleconferencing system is best described as a permanent Tele-Portal which permanently merges two common areas, giving remote participants a sense of virtual co-location while having a cup of coffee together, which will then automatically lead to the type of spontaneous, casual small talk which commonly emerges in coffee breaks.

  Besides the above listed systems and installations, numerous other Media Spaces have been created and studied since the late 1980s that are not included here (see Mackay (1999) for further reference). However, to close this section with the latest trends, two very recent projects that adopted and extended the idea of Media Spaces are explained in the following.

- *Mixed Reality Architecture (MRA):* Schnaedelbach et al. (2006) presented a system described as "Mixed Reality Architecture" (MRA) which allows the inhabitants of a building to share offices through permanent video links in the same way as in Media Spaces. As an extension, however, the docking between spaces is controlled across a shared three-dimensional virtual world interface, in which every office is represented as a box-shaped MRA Cell showing the video of the physical space it represents on its front side. These cells are mobile and can explore the virtual space in search for other cells. If two cells approach each other from front-to-front, at some point an audio-connection is established and the view is completely filled out by the video displayed on the cells front side. The offices are now shared as long as the cells remain connected in the virtual world. A long-term observational study in a real world setting could demonstrate the value of MRA for social interaction while remaining unobtrusive and allowing for privacy.

- *Pêle Mêle:* The provision of permanent video links between homes that offers family or friends several forms of communication ranging from casual awareness to highly focused two-way conversations at a distance is the goal of the video system “Pêle Mêle” (Gueddana and Roussel 2006). Pêle Mêle combines computer vision techniques with spatial and temporal filtering of video streams to display video based information about present or past activities. Awareness information such the availability of a remote participant are embedded in animations. For example, if a person becomes available for conversations, his or her video size automatically grows at the other end. Furthermore, computer vision techniques detect high activity levels which automatically cause significant actions to be recorded and replayed at the other end. This allows the system also to be used for asynchronous forms of communication.

2.4 Using virtual reality (VR) as a communication medium

This section outlines the recent emergence of a type of interface which mediates communication with others through the creation of shared virtual environments.

First, Section 2.4.1 explicates the concept of “presence”, which many researchers consider to be the “essence of any experience in a virtual environment” (Coelho et al. 2006). The section reviews common definitions of presence, lists contributing factors, and discusses common measurement approaches to assess a sense of presence. Then, Section 2.4.2 introduces the notion of “collaborative virtual environments (CVEs)”, which are special VR interfaces that allow several users to communicate with each other within a shared artificial spatial context. Section 2.4.3 goes into detail on the types of social interaction and awareness mechanisms that CVEs provide for communication and collaboration. Section 2.4.4 illustrates the concept of CVEs with several examples. Finally, Section 2.4.5 summarises empirical findings into human factors of CVEs.

2.4.1 The sense of presence in virtual environments

Virtual Reality (VR) refers to the real-time simulation of a three-dimensional, artificial, computer-generated world which is presented to a user in such a way that it appears as if he or she is actually "there". Users are typically able to explore a virtual environment from their first person perspective by actively moving and looking around. Special 3D-input devices such as data gloves allow for interaction with virtual objects. In addition to high fidelity graphics and surround sound, some advanced VR systems also provide olfactory or tactile (force) feedback to increase the realism of the virtual world as experienced by the user through all five senses.

In the beginning, research in VR was mainly centred around the technical feasibility of increasing realism and interactivity in single-user contexts. With the continuous maturation of VR systems, however, research interests turned towards the relationship between user and (virtual) environment. At the core of that relationship the concept of "presence" emerged and has been paid considerable attention to by researchers ever since. At a basic level, a user is "present" in a synthetic environment if that environment becomes his or her actual reality. Common definitions of presence include the "user's experience of *being there* in a mediated environment" (Held and Durlach 1992), "the (suspension of dis-) belief of being located in a world other than the physical one" (Slater and Usoh 1994), and the "perceptual illusion of non-mediation" (Lombard and Ditton 1997).

Presence researchers explore different types of presence, contributing factors, effects, and measurement approaches of presence. The following paragraphs give a brief summary of their main findings.

Contributing factors to a sense of presence: A sense of presence is a psychological phenomenon which is generated in the mind of the user. To understand what contributes to a sense of presence, both the technological components of a medium as well as the experience of the user have to be taken into consideration. IJsselsteijn and Riva (2003) therefore divide the determinants of the presence experience into *user characteristics* and *media characteristics*, which can be further sub-categorised into form factors and content factors. According to their general presence framework (see Figure 2.9), multi sensory stimulation arises from both the physical environment as well as the mediated environment and competes for the limited perceptual and attentional resources of the user. A medium that provides stimuli that are highly involving, immersive, realistic, or allow a high level of interactivity promotes a strong sense of presence, since more attentional resources are drawn towards the stimuli of the medium. Ideally, the user completely ignores the surrounding physical environment and feels part of the mediated virtual world. In this state, one "is there". Typical examples of presence enablers are immersive virtual environments (e.g. Cave Automatic Virtual Environment (CAVEs)), 3D computer games, IMAX movies, but also absorbing books, or even paintings.

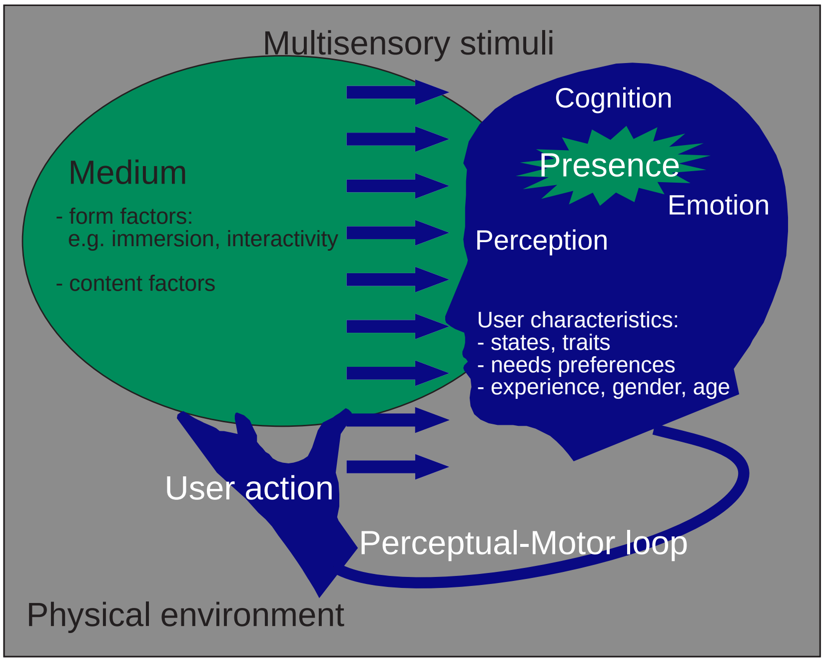


Figure 2.9: A general framework of presence, from IJsselsteijn and Riva (2003).

IJsselsteijn and Riva's (2003) general framework of presence also reflects other, empirically validated factor structures of earlier conceptions of presence, which distinguished between technology-specific variables such as immersion and realism on one side, and user-specific variables such as involvement, interactivity, and control on the other (see Witmer and Singer (1998) and Schubert et al. (2001)).

Physical presence and social presence: Based on existing conceptualisations of presence, IJsselsteijn and Riva (2003) distinguish between two types of presence: social presence (the feeling of being together) and physical presence (the sense of being physically located in mediated space). As previously outlined in Section 2.2.1, social presence is a characteristic of telecommunication media which depends on the richness of supported verbal and non-verbal communication cues. In contrast, a sense of physical presence depends more, as explained in the previous paragraph, on technological aspects and user characteristics.

IJsselsteijn et al. (2001) classified different media according to their support for one or the other type of presence (see Figure 2.10). Some media can be clearly assigned to one of the two categories: E-mails for example only convey social presence; single-user VR systems only convey a sense of physical presence. However, IJsselsteijn et al. also identify media that combine characteristics of social presence and physical presence which support a sense of being together in a shared space (here defined as a sense of co-presence). Media that fall within this category include Collaborative Virtual Environments (CVEs, explained in more detail in Section 2.4.2) and media utilising video, because they provide a mix of physical and social components.

User-studies produced empirical evidence for a positive correlation between physical presence, social presence, and co-presence in CVEs (e.g. in Slater et al. 2000, Nowak 2001). IJsselsteijn and Riva explain this interdependence by the fact that "there are likely to be a number of common determinants, such as the immediacy of the interaction, that are relevant to both social and physical presence". Yet, the exact interdependence between physical presence and social presence in telecommunication media remains a open research question.

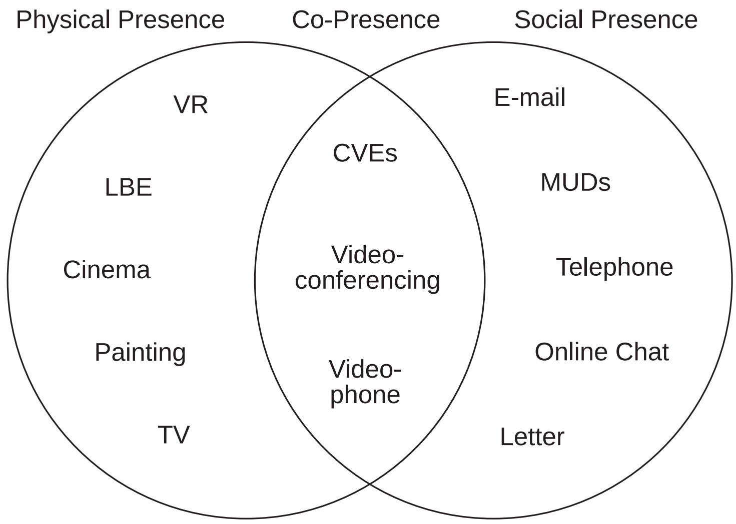


Figure 2.10: A graphic illustration of the relationship between physical presence, social presence, and co-presence with various media examples. Abbreviations: VR = Virtual Reality; LBE = Location-Based Entertainment; CVE = Collaborative Virtual Environment; MUDs = Multi-User Dungeons (adapted from IJsselsteijn et al. (2001)).

The effects of a sense of physical presence: If a user acknowledges a mediated, imaginary world to be his or her actual one, he or she responds to stimuli from that virtual world in the same way as he or she would in a real situation. Lombard and Ditton (1997) reviewed several physiological and psychological effects that were associated with a sense of presence. They list the most typical physiological effects to be arousal, motion-sickness, and automatic motor responses such as flinching, ducking, and tightly grasping one's chair. The most consistent psychological effects that are associated with a sense of presence are a high level of enjoyment and a high level of involvement with the presented content.

Many researchers also investigated the correlation between presence and task performance in tasks such as virtual training, spatial memory exercises, or searching tasks. Although several studies found physical presence and task performance to correlate, a clear causality between the two dimensions could not yet be clearly established and may vary with the individual task (see Nash et al. (2000) for a review).

Presence measures: The question of how to determine the user's sense of presence is one of the central issues in presence research today. In general, subjective, behavioural, and physiological methods are distinguished (Insko 2003).

Subjective measures rely on the participant's self-reported sense of presence. Users that have been exposed to a presence experience typically answer questions such as "How real did the environment seem to you?". Several validated presence questionnaires exist which assess a sense of presence through a combination of items targeting the underlying constructs such as immersion, involvement, realism, or control (see e.g. Witmer and Singer (1998), Lessiter et al. (2001), Schubert et al. (2001)).

Questionnaires are relatively easy to use and inexpensive to administer. They also allow the possibility of statistical analysis and do not interfere with the user's experience during the actual experiment. However, Insko (2003) also acknowledges some problems of this measurement approach: questionnaires are subject to response bias, are not able to measure the time varying qualities of presence, and might produce unstable and inconsistent responses depending on a participant's prior experience. Slater (2004) furthermore argues that subjects are mostly naive to the idea of presence and are therefore incapable of answering questions like "how present did you feel in the virtual environment?". According to Slater, responses to such direct questions measure a user's loose subjective associations with presence, which are likely to differ from the concept of physical presence in virtual environments experimenters actually seek to assess.

Behavioural measures are based on participants' responses that occur automatically, without conscious thought, such as ducking to avoid a virtual object that seems to fly at a immersed user (Sheridan 1992). While such measures give a strong indication for experienced presence, Insko (2003) points out that presence results obtained from video-analysis may be subjected to experimenter bias.

Physiological measures assess a sense of presence on the basis of physiological reactions by subjects. As an example, Insko (2001) reported fear behaviour responses, indicated by changes in heart rate and skin conductivity, to correlate with self-reported behavioural presence in a fear inducing virtual environment. Physiological measures involve complicated data acquisition and are mainly suited for stress or fear inducing environments that evoke strong and thus observable physiological responses.

Kramer et al. (2006) proposed an approach to measuring presence in computer-mediated communication based on linguistic features. They found that self-reported presence correlated with a number of task-independent linguistic features, including pronouns, terms referring to cognitive processes, and local deixis. People who felt more present in a remote environment made more use of local deixis (e.g. this, here) and the pronoun "we", and made less mention of cognitive and social processes.

Insko (2003) concludes a review of presence measures, that it is best to use "as many different measures as is feasible" at the same time when designing studies with a focus on presence.

2.4.2 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 (Benford et al. 2001). 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 which suggests him or her to be 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. Users 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.

2.4.3 The affordances of CVEs for remote collaboration

To describe the unique forms of collaboration and interaction that are supported in a CVE, this section outlines the affordances offered by CVEs in a similar way as has been done by Gaver (1992) with respect to media spaces. The concept of affordances originates in Gibson’s (1979) ecological approach to visual perception, where affordances refer to directly perceivable opportunities for interaction that an environment offers. With the essence of VR experiences being in the inclusive relationship between the participant and the VE, describing interactions within CVEs in terms of affordances is particularly appropriate. The affordances identified here are mainly derived from the *spatial model of interaction* (Benford and Fahlén 1993, Benford et al. 1994) as well as from a literature review of related work in the area of CSCW and VR (Gaver 1992, Benford et al. 1995, Gutwin 1997, Riva 1999, Fussell et al. 2000, Mantovani et al. 2006).

Affordances of a virtual space: CVEs describe a three-dimensional reference frame in which users are able to control and maintain their subjective views. This affords a sense of immersion, where the user acknowledges himself or herself as part of the virtual scene. A virtual space also affords the placement of more three-dimensional objects. These objects can be chosen carefully to constrain possible foci of attention which facilitates grounding processes. Virtual objects can furthermore be chosen and positioned to re-

Figure 2.11: An example of avatars collaborating in the "DIVE" collaborative virtual environment.

semble familiar arrangements of, for example, office furniture including round tables, interactive terminals, or shared displays such as those seen in Figure 2.11[7]. Virtual objects afford familiar interaction to CVE users. The spatial reference frame given by the VE furthermore affords the spatial integration of task space with interpersonal space.

Affordances of avatars: Where possible, avatars offer the same social affordances as humans. Avatars therefore are commonly designed as full-body anthropological models which can be tailored by users to match their characteristics like gender, age, and outfit (see the avatars in Figure 2.11). The appearance of an avatar therefore affords identification. Several avatars furthermore provide visual information about mood (happy face/sad face) and availability (eyes and ears closed/eyes and ears open) and may offer the user the ability to trigger a range of predefined expressive body animations such as waving hands or shrugging shoulders. These features afford non-

[7] Image source: www.sics.se/dive/demos/images/research.1.jpg, last accessed in December 2007.

verbal communication.

The representation of users and the projection of their actions into a Virtual Environment also brings about some other issues. The appearance of avatars as chosen by a user and as perceived by others does not have to match the actual characteristics of the user. Young, happy, pretty, female avatars may be controlled by old, sad, ugly men. This mind/body separation affords pretending and role playing on the acting side as well as scepticism and a reduction of truth worthiness on the perceiver side.

Affordances for vision: The three-dimensional nature of CVEs affords the deployment of stereoscopic and wide-field-of-view immersive displays such as head-mounted displays (HMD) or CAVE-like technologies. To the user, the opportunity of self controlled motion and gaze affords exploration and visual inspection of the VE.

Affordances for audio: The spatial relationship between different avatars as well as between an avatar and the VE affords a location-based modulation of sound to designers of CVEs. This modulation can be realised as directional surround sound as well proximity based volume adjustment. Spatial audio affords peripheral awareness for the user.

Affordances for interpersonal awareness: Avatars convey information about the presence and identity of a user. This affords background awareness by means of a quick glance to see who is present. Avatars also have a distinct front and back, indicating in which direction the avatar is "looking". This affords gaze awareness. The representation of virtual eyes on an avatar furthermore affords reciprocal views, mutual awareness and quasi-eye contact. Namely, if the eyes of another avatar are fully visible in the subjective view of a user, that user can infer that at the same time, the other participant sees a front view of his or her avatar. Gaze awareness and reciprocal views afford natural turn-taking as well as selective attention behaviour. Gaze awareness and reciprocal views furthermore afford the front wise approaching of an avatar if a participant wishes to raise its attention. The spatial relation between avatars also affords proximity behaviour, as inter-avatar distance can be perceived, negotiated, and adapted according to comfort levels. The spatial relations of other avatars can also inform about activity states and availability of others. Avatars that are standing in close proximity to each other in a face-to-face orientation are for example likely to be engaged in a conversation with each other whereas an avatar standing by himself will be more likely to be available and more appreciative when approached and engaged with. This affords socially appropriate behaviour.

Affordances for workspace awareness: Workspace awareness refers to knowing what others are working on in a shared workspace. CVEs afford workspace awareness in several ways. Shared workspaces are spatially integrated into CVEs which means that the proximity of an avatar to that workspace signals others that this avatar is engaged with the workspace. Closely linked with the subjective view of the workspace that can be inferred by others through the avatar's location and gaze direction, is the area in the workspace that is the focus for possible manipulations performed by the avatar. The proximity of an avatar can therefore inform about the authorship of a manipulation that occurs in the shared workspace. The mobility of the avatars furthermore affords smooth transitions between side-by-side or shoulder views towards the shared workspace and face-to-face views for example in the form of short glances that seek feedback. Virtual body extensions like virtual limbs, hands, or telepointers furthermore afford pointing at elements of the workspace, with the identity of the pointer being clear to others.

2.4.4 Examples of CVEs

CVEs were developed in various forms for purposes such as virtual business meetings, scientific co-visualisation, virtual therapy, and entertainment (Schroeder 1996, pages 1-2). The following paragraphs provide examples of CVEs within the classification of social online 3D environments and research systems.

Social online 3D environments: along with the explosion of the internet, several virtual environments emerged that follow the spatial interaction model. In these systems, graphical 3D environments or virtual worlds are displayed on a 3D browser application running on standard PCs. Users are embodied as a variety of avatars which are placed in persistent 3D virtual worlds where they can communicate, mingle, and play with others for the purpose of entertainment. Social online 3D environments can be further subclassified into massively multiplayer online games (MMOG) and online worlds for socialising.

MMOG such as *World of Warcraft*[8] or *Dungeons and Dragons*[9] foster the idea of role-playing within a given fictional plot and fantasy world. Groups of avatars either compete or cooperate in order to reach the goal defined by the game.

Online 3D worlds built for socialising such as *Active Worlds*[10], *There* [11], and *Second Life*[12], however, do not encourage role-playing, but rather allow users to portray themselves in a virtual world where they may interact with others, visit and build there own worlds, and even purchase or sell virtual objects and services. The idea behind creating a virtual world which is similar to the real world, but without physical limitations, has its roots in the vision of the *Metaverse*, as described in Stephenson's (1992) science fiction novel "Snow Crash".

Commercial social online 3D environments gained significant popularity and attention over the last five years and were able to create virtual communities counting millions of members distributed over the whole globe, with new members signing up every day.

Researchers recognised these systems as a valuable resource for studying social phenomena relevant to the field of CSCW such as the dynamics of establishing a virtual community (Hudson-Smith 2002), the design and use of 3D virtual worlds (Hansen 2002), forms of appropriation by users (Brown and Bell 2004), as well as the characteristics of collaborative play (Nardi and

[8] www.worldofwarcraft.com/, last accessed July 2007.

[9] www.wizards.com/default.asp?x=dnd/welcome, last accessed July 2007.

[10] www.activeworlds.com, last accessed July 2007.

[11] www.there.com/, last accessed July 2007.

[12] www.secondlife.com/, last accessed July 2007.

Harris 2006).

Research systems: Experimental CVE platforms were developed in various computer science labs since the mid-1990s aiming to explore the design space of this new technology. Trial systems and prototypes often made use of complex and cutting-edge technology such as high-end computer graphics hardware and immersive stereo displays, as well as specialised input and output devices. In the following, two representative CVE research platforms are described in more detail.

- *MASSIVE:* The research initiative MASSIVE (Model, Architecture, and System for Spatial Interaction in Virtual Environments) was an early attempt to develop a CVE for teleconferencing (Greenhalgh and Benford 1995). MASSIVE realised several scenarios that were inspired by the spatial model of interaction that Benford et al. (1994) had developed earlier. As such, MASSIVE implemented forms of negotiating awareness through the use of *aura* (a subspace within which an object interacts), *focus* (the more an object is within your focus, the more aware you are of it), and *nimbus* (the more an object is within your nimbus, the more it is aware of you) and demonstrated the basic functionality of first, crude prototypes in distributed settings. MASSIVE also served as a test bed for later user tests (e.g. Bowers et al. 1996) and was eventually developed further in the system "MASSIVE-2" which allowed for the integration of further contextual factors for improved awareness (Benford and Greenhalgh 1997).

- *COVEN:* The goal of the European COVEN (COllaborative Virtual ENvironments) project was to develop a comprehensive understanding of the nature of CVEs, and to develop a sophisticated platform for next-generation CVEs (Normand et al. 1999, Frécon et al. 2001). The project lasted for four years and involved twelve academic and industry partners. CVE prototypes were developed based on the DIVE (Distributed Interactive Virtual Environment) toolkit (Carlsson and Hagsand 1993, Frécon and Stenius 1998), which was improved and extended in the course of the project. The main improvements that the

COVEN project brought forward include humanoid avatars featuring a dynamic level of detail and motion control, a streamlined networking data infrastructure, and the modular integration of spatial sound as well as video. In addition to these technological achievements, usability factors of several application scenarios including a virtual conference, interior arrangement, and travel rehearsal were explored.

As part of the COVEN initiative, a new visualization technique, "subjective views", was also introduced as proposed earlier by Smith (1996) which allows two participants who share the same VE to tailor their view of shared objects to highlight or de-emphasise visual features according to their individual needs. A virtual table could thus appear as solid geometry for one user while the same table might appear as a wire frame for the other user.

To allow rapid development of customised CVE prototypes within the COVEN platform, Frécon and Smith (1999) developed the high-level behaviour language DIVE/TCL, which allows researchers easy access to otherwise complex state-of-the-art CVE technology.

2.4.5 Empirical research in CVEs

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 2.10 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. (1996) 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 a human-like avatar compared to a control conditions that embodied interlocutors as talking dogs. However, Nowak and Biocca (2003) found people preferred a less anthropomorphic rep-

Focus of study	References
Avatar appearance	Parise et al. (1996) Nowak and Biocca (2003) Garau et al. (2003) Bailenson et al. (2005)
Usability inspection	Greenhalgh and Benford (1995) Normand et al. (1999)
Turn-taking behaviour	Bowers et al. (1996)
Small group dynamics	Slater et al. (2000)
Cross-media comparisons	Nakanishi et al. (1998) Sallnäs (2005)

Table 2.10: User studies in CVE-research.

resentation 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. (2003) and Bailenson et al. (2005) also explored the relation of pictorial realism (how real avatars look) and behavioural realism (how human-like avatars behave). Garau et al. (2003) 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 sum, 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 (1995) and Normand et al. (1999) 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 to form a circle with their avatars.

Bowers et al. (1996) 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 wish to interact with.

Slater et al. (2000) 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 two CVE-studies, the author is aware of, involved comparisons of VMC and avatar-mediated communication. Nakanishi et al. (1998) 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 nonmediated 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 copresent.

The second example is a series of two studies reported by Sallnäs (2005) 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 mostly exploratory nature of most of the studies as well as the heterogeneity of the measures that were used.

2.5 Chapter summary

This chapter was split into four main sections which introduced the background of this research work. Section 2.1 established the theoretical understanding of interpersonal communication and collaboration. Building on Shannon and Weaver's (1949) communication model, collaborative interpersonal communication can be described in terms of the exchange of verbal and non-verbal messages to establish and develop a shared understanding of the task at hand.

Section 2.2 highlighted how interpersonal communication is affected by the use of different telecommunication technologies. Different media were classified with regards to their richness, their support for grounding mechanisms, their social presence, and with regards to the level to which they support a sense of "being together". In all classifications, media that support more verbal and non-verbal channels are more similar to face-to-face. People's media selection is mainly driven by the type of collaborative task they want to use a medium for. One particular challenge for mediating collaborative activities with groupware systems that involve a shared workspace is supporting workspace awareness, that is, to constantly inform remote participants about the others' activities in the shared workspace.

Section 2.3 focused on video-mediated communication. It included a comprehensive review of cross-media studies that compared audio-only, audio-video, and face-to-face collaboration. The review pursued two goals. First, it gave an overview of the range of experimental tasks and measures that are typically applied in cross-media studies. Second, the summarised findings demonstrated our current understanding of the benefits and limitations of video-mediated communication. The second half of this section then reviewed research approaches that attempt to improve video-mediated communication.

Finally, Section 2.4 explained the concept of Collaborative Virtual Environments (CVEs) as a new communications medium and illustrated their potential to support remote collaboration. This section also included a more detailed review of the concept of "presence", which is the key user experience dimension in virtual environments. Research into human factors of CVEs is still in its infancy. User studies that have been conducted were often exploratory usability inspections, or focused on the impact of different avatar representations on social responses. Only a few experiments have compared avatar-mediated communication in CVEs with video-mediated communication leaving several research questions into the value of spatial virtual interaction unanswered.

Based on the foundations laid by this chapter, the following chapter explains the research approach pursued in this thesis to address some of these questions.

Chapter 3

Research Approach

The previous chapter gave an overview of the context and the state of the art in VMC research, including underlying communication models, main empirical findings, methodological challenges, as well as new interface directions that have been explored. One example of these new interface directions is the approach of Collaborative Virtual Environments (CVEs) (introduced in Section 2.4). The goal of this research is to enhance our understanding of video-CVEs as a medium to support real-time distributed work. Collaboration in video-CVEs differs substantially from normal video-collaboration. What opportunities and challenges does spatial virtual collaboration with video-CVEs entail? The next four sections explain how this thesis work seeks to find to an answer to this question.

First, Section 3.1 introduces the concept of live video embodiments in video-CVEs. Then, it presents the video-CVE "cAR/PE!" in more detail as cAR/PE! serves as the research platform for this work. Section 3.2 gives an overview of the research method. It states the main research hypothesis and presents the strategy chosen to investigate and test it. Finally, Section 3.3 gives a brief overview of the four experiments conducted.

3.1 A CVE approach to video-conferencing

Avatars – user embodiments within a CVE – can be implemented in many different ways ranging from simple geometric shapes to fantasy characters to realistic, expressive full-body virtual models (see Figure 3.1).

A mixed-reality approach to embodiments in virtual environments represents users directly by live-video images which are texture-mapped onto virtual navigable planes. CVEs that follow such a video-avatar approach

(a) Avatar in "Active Worlds" (b) Avatar in "Second Life"

Figure 3.1: Examples of avatars in online 3D worlds.

include Freewalk (Nakanishi et al. 1996), CU-SeeMe (Han and Smith 1996), Gaze-2 (Vertegaal et al. 2003)(see Figure 3.2), Coliseum (Baker et al. 2003), and cAR/PE! (Regenbrecht et al. 2004). The following section describes the system cAR/PE! in more detail because it serves as the video-CVE platform for this research.

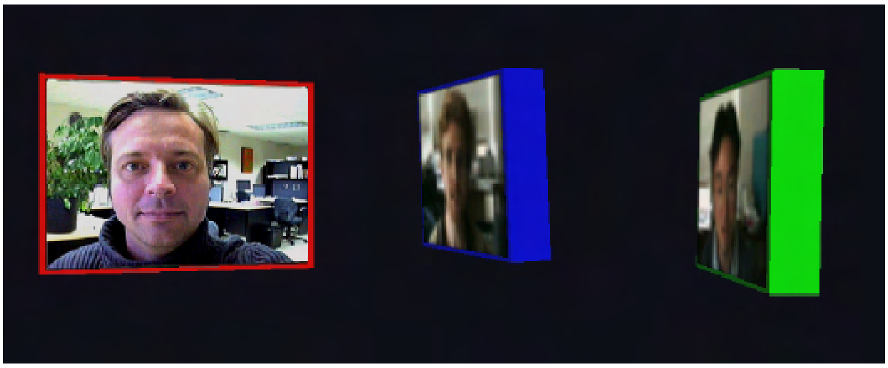


Figure 3.2: Video-avatars in Vertegaal et al.'s (2003) system "Gaze-2".

3.1.1 The video-CVE "cAR/PE!"

The cAR/PE! (collaborative Augmented Reality for Presentation and Engineering) (Regenbrecht et al. 2004) prototype was developed by the Virtual Reality Competence Center (VRCC) at the DaimlerChrysler Research and Technology Centre in Ulm, Germany. The main motivation behind the development of cAR/PE! was the integration of a shared workspace with video-conferencing which would allow distributed expert teams to collaborate efficiently around various forms of data. As such, cAR/PE! is an attempt to provide distributed teams the best possible direct *simulation* of a face-to-face meeting.

The cAR/PE! conferencing space: The virtual world in cAR/PE! consists of a conference room which includes elements commonly found in real meeting situations. The rationale behind the use of a room metaphor was an attempt to afford the same kinds of behaviours that occur in real world meetings, and thus allow remote participants to seamlessly continue their established work practices. Figure 3.3 shows a wide-angle view of the cAR/PE! room with the main elemets labelled.

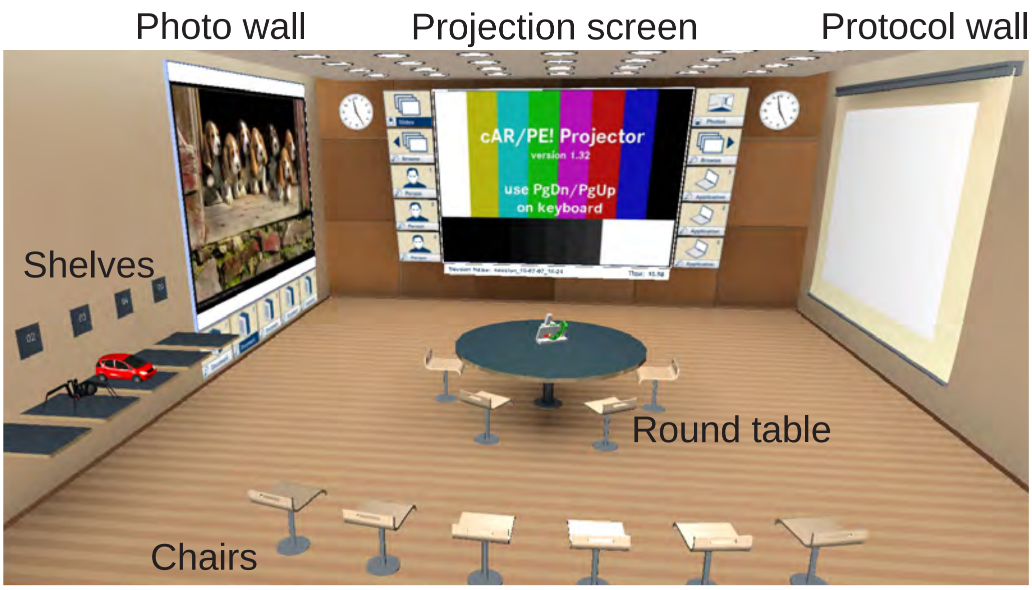


Figure 3.3: The cAR/PE! virtual conferencing room.

There are several interactive elements within the cAR/PE! room. Chairs mark significant positions for avatars and provide an obvious place to gather around the table located in the middle of the room. Users can also pick up 3D-models from shelves on the wall and place them on the centre of the table where they can be inspected and discussed. On one wall, there is a projection screen which serves as a shared display where users can present slides to each other or share applications. Other walls integrate interactive areas which allow users to upload a set of images or take minutes during a meeting.

Video embodiments: Participants in a virtual meeting in the cAR/PE! environment are embodied by their video avatars. Figure 3.4 shows a video avatar and illustrates how its orientation follows the user's actual view into the virtual room. The front of the avatar shows a video-texture of the user, while the back face is rendered transparently (Figure 3.4 on the right) to reduce occlusion in situations where other avatars look in the same direction.

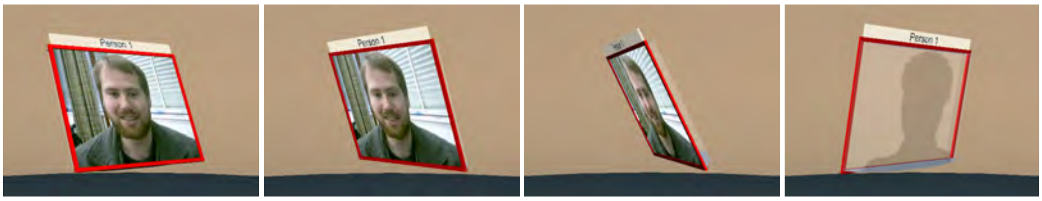


Figure 3.4: A video avatar in cAR/PE! changes its orientation according to the view controlled by the user.

The spatial alignment of a user's video avatar supports gaze awareness for the other users. However, the gaze direction of video avatars is not "spatially faithful" as defined by Nguyen and Canny (2005). That means that other users can only *infer* from the orientation of the video-plane what the person in the video can see, rather than being able to directly *perceive* the gaze of a person as it appears in the video. This is firstly caused by the offset between camera and monitor which causes the person in the video not to look *straight* out of the video plane. Secondly, even if this offset is minimised (for example by using a video tunnel), the video would suggest eye-contact at all times even in cases where the video is highly distorted due to the orientation of the avatar (see Figure 3.5). This effect is called the *Mona Lisa effect* named after the famous painting by Leonardo Da Vinci showing the portrait of a woman whose enigmatic gaze meets the viewer's gaze, regardless of which angle she is observed from.

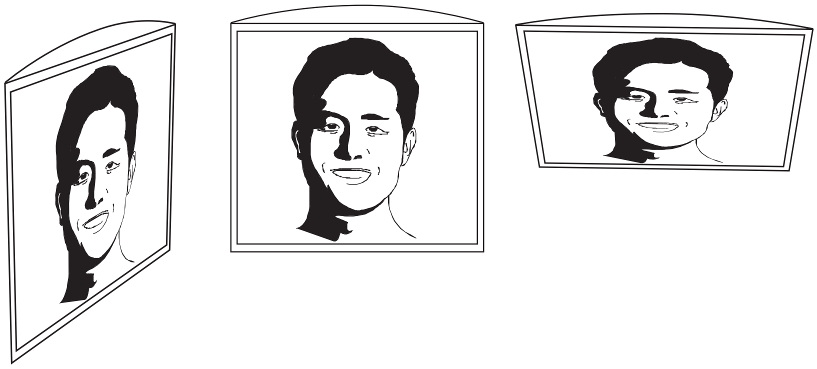


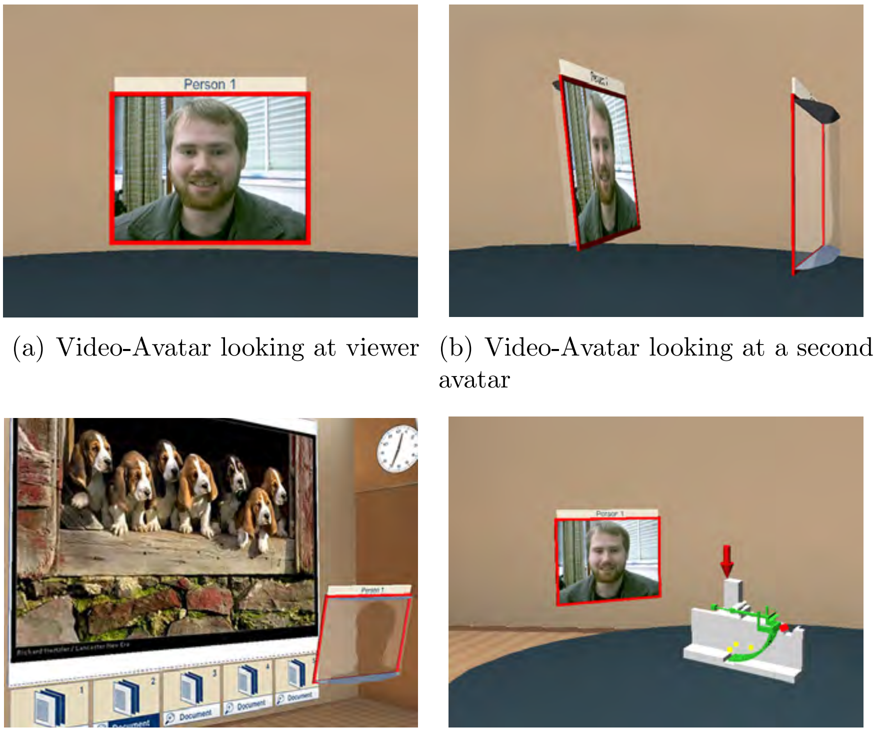
Figure 3.5: The Mona Lisa effect: the person in all three video-avatars appears to look straight at the viewer. However, only the video-avatar in the middle is actually oriented towards the viewer. The *perceived* gaze direction of the person in the video and the actual gaze direction that has to be *inferred* from the avatar's orientation are conflicting in the video-avatar on the left and on the right.

Vertegaal et al. (2003) and Baker et al. (2003) proposed multi-camera approaches to overcome these limitations. However, none of them are implemented in the cAR/PE! prototype to date.

Despite these limitations, the four examples shown in Figure 3.6 illustrate how video-avatars can convey important awareness information.

Figure 3.6(a) shows the front of a video-avatar of a person within close proximity of the viewer. From such a spatial arrangement the viewer can infer that his or her avatar must also be in the focus of attention of the other avatar, who might want to engage in a conversation. This spatial arrangement therefore serves as the equivalent for reciprocal eye-to-eye contact.

In Figure 3.6(b), the viewer is watching two other avatars which are facing each other closely. This shows that these two people are likely to be talking with each other and may only be approached with caution in order not to disturb their conversation.



(a) Video-Avatar looking at viewer (b) Video-Avatar looking at a second avatar

(c) Video-Avatar looking at the photo on the wall (d) Video-Avatar looking and pointing at the model on the virtual table

Figure 3.6: Examples for gaze awareness in cAR/PE!.

The position of an avatar within the room can also provide awareness about activities of the participants. For example, in Figure 3.6(c), the avatar is looking at the photo displayed on the photo wall, which can be inferred by the viewer from the avatar's close proximity to and orientation towards the photo wall.

Finally, Figure 3.6(d) shows an avatar on the other side of a virtual table on which a 3D model is placed. In this case, the viewer can not only infer that the other person is currently looking at the model, but also from what angle he is looking at it. In addition, the avatar is pointing at a part of the model, indicated by the telepointer.

Implementation: The cAR/PE! application is based on the DaimlerChrysler in-house VR package "DBView" (Sauer 2001), which provides several components and modules that can be flexibly combined and thus allow fast prototype development. The heart of a VR-application is an encapsulated scene-graph which contains the information about geometry and appearance of all virtual objects of a virtual world. All scene graphs in DBView are implemented with Open Inventor[1], an OpenGL based scene graph API for facilitating graphics programming. An object management system handles the interaction with elements of the scene graph. Furthermore, DBView features a module management system as well as a messaging component which inter-connects the modules used within an application. To accommodate cAR/PE!'s particular requirements for remote collaboration, specific modules for video transmission, spatial audio, 3D display and application control were implemented.

The cAR/PE! application runs locally on standard PCs, supporting up to six users who are connected through an Ethernet network. Between the users, information about the actual scene-graph is semi-replicated. That means that information about the initially loaded scene is accessed locally on every PC during start-up. In order to guarantee consistency, participants therefore need to distribute the room geometry, 3D models, and presentation slides that will be used beforehand. Lauwers et al. (1990) and Dewan (1999) give a review of differences between centralised and replicated architectures for collaborative applications.

Figure 3.7 shows the two different ways cAR/PE! stations connect with each other. Firstly, audio, video, and data exchange is enabled through direct point-to-point connections between all participating parties. Secondly, messages that include information about interaction within the cAR/PE! room such as position updates of video avatars or data manipulation events are distributed to all participating parties via a centralised "Common Request Broker".

cAR/PE! is a technically advanced but early proof-of-concept research prototype which can be considered as the starting point for an iterative design process. An effective improvement presupposes a well-founded knowledge of the opportunities and issues of spatial virtual interaction. Human factors of

[1] http://oss.sgi.com/projects/inventor/ , last accessed in July 2007.

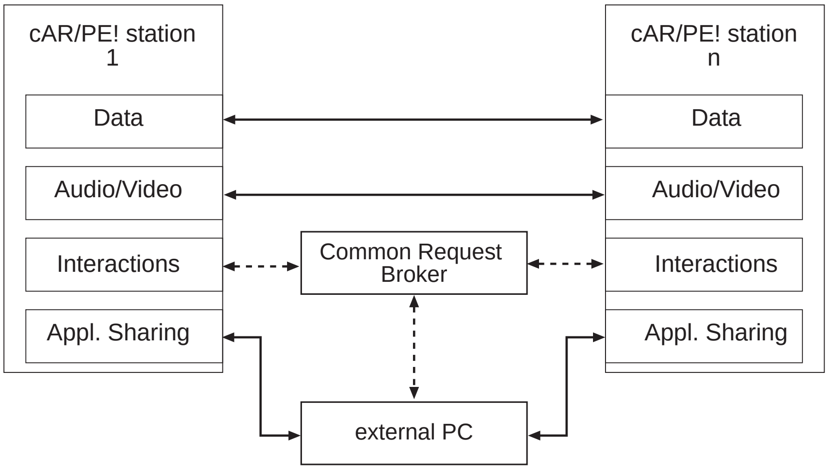


Figure 3.7: The cAR/PE! network architecture (adapted from Regenbrecht et al. (2006)).

CVEs, however, are poorly understood (Benford et al. 2001).

The main goal of this research is to address this shortcoming by enhancing our understanding of human factors of video-CVEs.

3.2 Method

The following sections explain the method applied in this research. The main hypothesis of this thesis is that geographically dispersed people who collaborate in video-CVEs feel and behave more similarly to being face-to-face than if they collaborate through standard video-conferencing systems.

The next sections will outline the details and considerations that influenced the design of four user studies that investigated and tested this hypothesis. First, Section 3.2.1 gives the rationale for investigating human factors of video-CVEs interaction by means of contrasting them with standard video-conferencing as well as face-to-face situations. Then, Section 3.2.2 presents the selected independent measures, i.e. interface design dimensions that allow a systematic and quantifiable distinction between video-CVEs and standard video-conferencing systems. Section 3.2.3 lists the dependent measures applied to discriminate the extent to which participants "feel like they were face-to-face" and the extent to which participants "behave like they were face-to-face" in different media conditions.

3.2.1 A comparative approach

The core of this research is a series of comparative user studies that involve video-CVEs, conventional video-conferencing tools, face-to-face situations and further control conditions. This comparative approach was chosen for the following four reasons:

First, despite the ongoing discussion of the value of adding video to audio, conventional video conferencing systems are the most common face-to-face replacement tools today. Contrasting cAR/PE! with them therefore allows us to identify existing differences to today's "state-of-the-art". Including an unmediated face-to-face condition as a "gold standard" reference allows us to further qualify and interpret these differences. Taken together, the results are thus able to show where CVEs are, and are not, able to provide a user experience that is closer to face-to-face collaboration than is possible with current video-conferencing tools.

Second, users in cAR/PE! are embodied in a very similar way as they are in other video-conferencing applications – namely by their real-time video. That way, empirical comparisons of cAR/PE! with standard video conferencing tools do not have to take the varying social responses associated with different avatar representations into account. Thus other (so far neglected) design dimensions can be investigated or extracted without the interference of unwanted noise caused by different embodiments.

Third, a substantial knowledge gap can be addressed, since only very few cross-media studies that include CVE as well as video-conferencing conditions can be found in literature to date (see Section 2.4.5).

Fourth, this research is subdivided into several interrelated comparative studies which have some measures, types of tasks, and conditions in common. That way, methodology and results can not only be compared within but also, to some degree, between the individual experiments. This is rare in the often controversial VMC literature and allows insights with high external validity.

3.2.2 Independent measures

CVEs and standard video conferencing tools differ in their nature. How can these differences be systematically characterised, quantified, and translated into experiment conditions? This section explains the systematic approach this research pursued in order to arrive at an answer to these questions.

A useful interface design space that accommodates the most fundamental differences between standard video-conferencing systems and video-CVEs is given by Benford et al.'s (1996) taxonomy of shared spaces. The following paragraphs introduce its underlying dimensions of *spatiality*, *transportation*, and *artificiality*, and illustrate their relevance for the research scope of this thesis.

Spatiality: The dimension of spatiality concerns a shared space's

> "level of support for fundamental physical spatial properties such as containment, topology, distance, orientation, and movement, ... Its extremes are characterised by the notion of place, containing context for participants; and space, a context that further provides a consistent, navigable, and shared spatial frame of reference" (Benford et al. 1996).

Regular video-conferencing interfaces provide no shared internal structure, no dimensions or controls, and therefore little spatiality, other than containment. Regular video-conferencing therefore is a *place*.

In contrast, CVEs re-introduce a shared three-dimensional reference frame and are thus examples of a collaborative *space*. The concept of a consistent space also emerges if the videos of remote people are presented in a spatially consistent way as done in "Tele-portals" (see Section 2.3.2). Figure 3.8 shows the dimension of spatiality with the examples of regular video conferencing, Tele-portals, and CVEs.

Distinguishing collaborative and communicative environments in terms of the concepts of "place" and "space" has proven to be useful within the CSCW-community (Harrison and Dourish 1996, Dourish 2006). Comparing regular video-conferencing tools and video-CVEs with regards to these concepts can therefore be regarded as well-founded and theoretically relevant for the still ongoing "place-versus-space" debate.

Figure 3.8: Benford et al.'s (1996) dimension of spatiality and examples.

Transportation: The dimension of transportation is closely related to the virtual reality concept of immersion and concerns the

> "extent to which a group of participants and objects leave behind their local space and enter into some new remote space in order to meet with others, versus the extent to which they remain in their space and the remote participants and objects are brought to them. It therefore characterises the essential difference between the concept of local and remote" (Benford et al. 1996).

Unmediated face-to-face meetings take place entirely in a local space. Similarly, when people meet via conventional video conferencing, that meeting can still be conceived to be mostly local, since participants remain in their local space while information about remote partners and objects are brought to them.

A CVE on the contrary, is an example of a *remote* meeting, since participants are immersed into and involved with a virtual, remote space where they can meet others. The level of immersion supported by different displays is furthermore correlated with the remoteness of a CVE. Fully immersive displays such as head-mounted displays or CAVEs are able to cut out any visual stimulus from a participant's local environment and thus are considered to be more remote than, for example, small screen desktop-CVEs which cannot exclude stimuli and distractions from a participant's local environment. Figure 3.9 shows the dimension of transportation with examples of face-to-face meetings, regular video conferencing, and display-dependent CVEs.

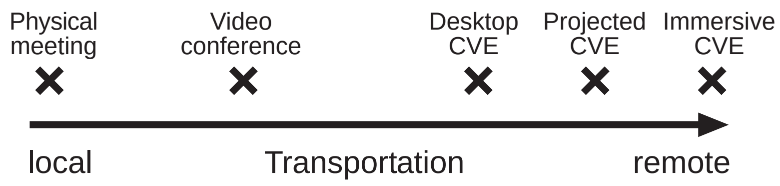


Figure 3.9: Benford et al.'s (1996) dimension of transportation and examples.

At the heart of any comparison of collaborative systems with respect to the degree to which they "transport" a user to a remote space lies the fundamental relationship between the concept of *presence* and its impact on social interactions. Investigating the nature of this relationship bears both theoretical and practical importance. Firstly, results of comparative studies may inform about the psychological consequences of the presence phenomenon in the VR sub-domain of CVEs. Secondly, results of these comparisons may also inform future CVE designers if the investment in expensive immersive technologies can be expected to lead to a significant improvement in the collaborative user experience.

Artificiality: Benford et al.'s (1996) taxonomy for shared spaces includes a third dimension – artificiality. According to a system's ratio of real-world to computer-generated information present it can be classified to be either "physical" or "synthetic". CVEs with computer generated avatars are fully synthetic, while a video-phone would be entirely physical. Figure 3.10 shows the dimension of artificiality with different versions of video conferencing and CVEs.

Since video-CVEs mix real-world (videos) with artificial content (VE), they are both synthetic and physical. Similarly, video conferencing tools that include a shared workspace are both physical and synthetic. With this common hybrid characteristic, comparisons between these kinds of systems with respect to their degree of artificiality can therefore not be expected to produce useful results.

Consequently, the research design space for this work was limited to the dimensions of spatiality and transportation.

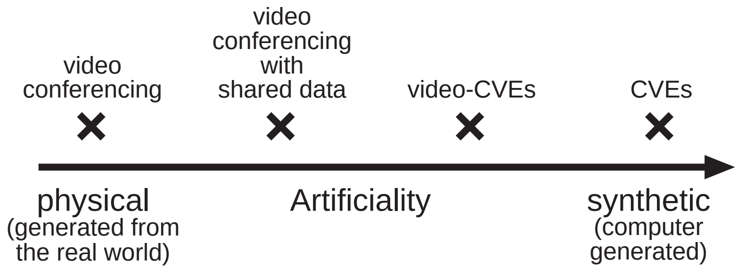


Figure 3.10: Benford et al.'s (1996) dimension of artificiality and examples.

3.2.3 Dependent measures

To test the main research hypothesis, subjective measures were needed that are able to assess and discriminate the extent to which participants "feel" like they were face-to-face, and objective measures were needed that are able to assess and discriminate the extent to which participants "behave" like they were face-to-face in different media conditions.

Several subjective cross-media metrics that see face-to-face communication as the gold standard compared to mediated communication exist, but have not been applied in comparative studies involving video-CVEs and standard video-conferencing conditions. It was therefore not clear which of them would be the most suitable and sensitive enough for the focus of this research.

The pragmatic approach taken was therefore to trial multiple subjective measures for their capability to discriminate between video-CVEs and standard video-conferencing first, and then to focus on and reapply the measures that can be expected to produce the most useful results. As such, two social presence scales, one physical presence scale, and a self-designed questionnaire targeting subjective communication quality were explored for their applicability in the first two studies, out of which the well-proven ones were reapplied in later experiments and complemented with questionnaires targeting copresence, awareness, ease of use, and preference to suit the individual research question at hand.

The degree to which participants behaved like they were face-to-face while using the video-conferencing systems was investigated by analysis of recorded videos and of communication transcripts. These analyses assessed and compared aspects of collaboration including verbal effort, distraction, view coordination, as well as an in-depth content analysis of conversation transcripts based on word counts.

The measures applied focus primarily on the process and satisfaction of tele-collaboration, since these measures are more likely to produce differences than measures focussing on the outcome of collaboration (see Section 2.3.1).

3.3 Overview of the experiments

The research hypothesis was investigated and tested in four related experiments which all addressed different underlying questions. This section gives an overview of the motivation and goals of each study, and explains the relationships between them and the conditions involved.

	Study 1	Study 2	Study 3	Study 4
Name	Desert Survival Game (DSG)	Dream House	Dogs & Owners	Celebrity Quotes
Details in	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Task type	Conversational game	3D design review	Photo review	Collaborative game
Main objective	explore social presence measures	replicate and strengthen findings	investigate impact of spatiality	investigate impact of transportation

Table 3.1: The four experiments in chronological order.

Table 3.1 shows the four studies in chronological order and names them according to their experimental tasks. These names will be used consistently in the remainder of this thesis to refer to the studies. As can be seen in the table, tasks and task types varied across the four studies.

The studies built upon each other. Experiment "DSG" explored suitable subjective measures which allow insights as to which VMC interface supports a "feeling" more similar to being face-to-face. It tested two social presence scales and a questionnaire on communication aspects for their power to discriminate between a standard video-conferencing system, a video-CVE, and a face-to-face condition. Results obtained from each scale were also explored for indications whether video-CVEs push the collaborative experience further towards the face-to-face gold standard.

Experiment “Dream House” re-applied one of the two social presence measures in an attempt to replicate and strengthen a finding from the first experiment. It also explored the questions if a higher level of transportation of an interface leads to a higher sense physical presence at a remote space, and, whether or not there is a relationship between the sense of physical presence and social presence.

The findings of the first two studies provided valuable experiences, both methodologically and substantively, which informed the design and the more specific research questions of experiments “Dogs and Owners” and “Celebrity Quotes”.

The study “Dogs and Owners” investigated if added spatiality in VMC fosters face-to-face-like collaboration. It applied a set of subjective scales and included an objective video analysis to compare collaborative behaviour.

Finally, the fourth experiment “Celebrity Quotes” complements the third study by changing the focus to the transportation aspect of a video-CVE. In particular, it investigated if video-CVEs with a higher level of transportation induce a feeling and support collaboration that is more similar to being face-to-face. It used subjective measures, video analysis, as well a relatively new method of analysing communication transcripts based on word counts.

Table 3.2 summarises the types of measures applied by each experiment in a side-by-side view.

	Subjective ratings	Video analysis	Transcription analysis
DSG:	•		
Dream House:	•		
Dogs and Owners:	•	•	
Celebrity Quotes:	•	•	•

Table 3.2: Types of measures applied in the four studies

Experimental Conditions: Five different experimental conditions were applied and some of them were used repeatedly. For example, all studies included a standard video conferencing condition and a desktop video-CVE. Further spatial control conditions were added in compliance with the given research focus. All conditions were furthermore designed and coordinated to cover different points along the dimension of transportation (see Table 3.3).

The experimental conditions can also be categorised based on their support for fundamental spatial properties into *places* or *spaces*. Table 3.4 shows one such classification. The face-to-face conditions are not included here since the concept of spatiality applies only to interface properties of shared-space systems (Benford et al. 1996).

	FtF meeting	Video table	2D video-conf.	Desktop video-CVE	Immersive video-CVE
DSG:	•		•	•	
Dream House:			•	•	
Dogs and Owners:	•	•	•	•	
Celebrity Quotes:	•		•	•	•
	local		Transportation →		remote

Table 3.3: The conditions included in the four experiments as indicated by a bullet point, arranged along the dimension of transportation.

	2D video conferencing	Video-CVE	Video table
DSG:	•	•	
Dream House:	•	•	
Dogs and Owners:	•	•	•
Celebrity Quotes:	•	•[a] •[b]	
	"Place"	"Space"	

[a]desktop monitor
[b]stereoscopic projection

Table 3.4: Conditions included in the four experiments as indicated by a bullet point, categorised according to their degree of spatiality as *place or space.*

As mentioned, the study "Dogs and Owners" focused on the impact of added spatiality in VMC systems. Besides a standard video conferencing and a video-CVE condition it also included the control condition "Video table", an interface based on video streams in a physically-fixed arrangement around an interactive table. The rationale behind this approach is that by comparing two different spatial interfaces with the non-spatial one, differences that are found can be more easily ascribed to high level spatial properties rather than to low level usability features of the systems involved. Since the condition "Video table" creates a shared reference frame based around a local, physical table, it is located closer towards the local end on the transportation axis.

The study "Celebrity Quotes" investigated the dimension of transportation, or more specifically, the impact of more or less "remoteness" of an interface. It included the control condition "Immersive video-CVE", which utilised a large-screen stereo-projection to display the same virtual environment which was otherwise presented on a smaller, conventional desktop screen. The rationale for this control condition is that bigger and more immersive displays are more likely to induce a stronger sense of actually being or existing in the displayed virtual, remote space by blocking out more stimuli from one's actual, local environment. On the transportation axis the condition "Immersive video-CVE" therefore assumes a position which is more remote than the condition "Desktop video-CVE".

In all conditions, the dimensions of transportation and the notions of space and place as extremes of the dimension spatiality are related. Spaces (interfaces that support distant users with a consistent, three-dimensional, shared, navigable environment) can only be either local (remote information is consistently integrated into the user's local space) or remote (user leaves body behind and enters in a shared virtual space). Places, on the contrary, only contain a non-spatial context for participants and hence are neither local nor remote. On the transportation axis, places therefore always lie between local and remote spaces.

Participants: The subjects for the user studies were recruited from students and staff at the University of Canterbury and the University of Otago. Either two or three participants formed a group which first used and then evaluated the experimental conditions. The more explorative studies "DSG" and “Dream House” did not assume specific participant profiles in the recruitment process but investigated demographic effects on the data collected to inform the recruitment process of subsequent experiments.

The studies “Dogs and Owners” and “Celebrity Quotes”, however, had a more specific research focus. To minimise unwanted side effects participating teams were recruited as “same gender friends”. Table 3.5 shows a side-by-side overview of participant parameters involved in the studies.

	Participants	Group size	Gender	knew each other before?
DSG:	42	3	*	*
Dream House:	36	2	*	*
Dogs and Owners:	26	2	same sex	yes
Celebrity Quotes:	36	2	same sex	yes

* = not controlled

Table 3.5: Overview of participants recruited in the four experiments.

The following chapters present each individual study including the results obtained in more detail.

Chapter 4

Experiment: "Desert Survival Game"

The overarching goal of the four experiments described in this thesis was to investigate if people collaborating in video-CVEs feel and behave more similarly to being face-to-face than people collaborating through standard video-conferencing tools. Investigating human factors of a communications medium by contrasting it with other media and face-to-face conditions is a common practice in CSCW research (see Section 2.3.1). However, as mentioned in Section 3.2.1, there is almost no research literature on cross-media comparisons which included standard video-conferencing conditions and video-CVEs. Consequently, there is no established set of measures or heuristics that specifically apply for media comparisons involving standard video-conferencing conditions and video-CVEs.

The main goal of the "Desert Survival Game" experiment described in this chapter was to explore the utility of several subjective rating scales for comparisons between standard video-conferencing systems and video-CVEs, allowing conclusions to be drawn about which of them afforded a user experience that was more similar to being face-to-face. This study therefore provided valuable experience for later experiments, both with regards to the selection of measures and more specific research questions to be asked. The subjective rating scales under investigation were

- a set of bipolar pairs to assess social presence, defined by Short et al. (1976) as a "subjective quality of the medium".
- a questionnaire to measure social presence, defined by Biocca et al. (2001) in terms of the "experience of the other persons".

- an exploratory questionnaire including items on subjective communication and task performance.

Short et al. (1976) conceive social presence as "the degree of salience of the other person in the interaction" which is directly related to non-verbal channels available in a communications medium (see Section 2.2.1). Social presence, following this definition, is therefore a subjective quality of the medium itself, which a user perceives and is able to externalise in form of a set of attitudinal scores. Biocca et al. (2001) conceive social presence as a psychological state, which is modulated by a sense of co-presence, psychological involvement, and behavioural engagement (see Section 2.2.1). Although carrying the same label, social presence therefore circumscribes different aspects of the user experience of mediated communication. Nevertheless, both of these conceptualisations of social presence see face-to-face as the gold standard and are thus of interest for this study. To avoid confusion between the two types of social presence, this chapter explicitly differentiates between social presence as a perceived subjective quality *of* a medium, and social presence as the experience of the other *in* a collaborative situation. In line with the research hypothesis it was expected that the video-CVE would yield a higher social presence than the standard video-conferencing condition. A further set of questions targeting subjective communication quality and task performance was also trialled and explored for its applicability for purposes in this research. These measures were applied in a direct comparison between a normal video-conferencing interface, a video-CVE, and a face-to-face reference condition.

Another goal of this experiment was to gather information on demographic factors to be considered during the recruitment process for following experiments. Therefore, the data obtained was examined for potential interactions between media conditions and the age group of participants, and between media conditions and the simulator experience of participants.

Finally, the experiment sought to collect first experiences of emerging usage patterns and possible usability issues that surface in the video-CVE condition, which could be used as specific questions in following experiments.

4.1 Experiment design

The study was designed and conducted as a joint effort between the author and members of the Multi-Media Systems Research Laboratory (MSRL) at the University of Otago.

4.1.1 Experimental task

The experiment focused on three-party conversations. As an appropriate task within that focus, it applied the "Desert Survival Game" (Lafferty et al. 1974), a common team-building communication exercise, in which members of a group are placed in the imaginary extreme situation of having crash-landed their plane in a desert, and now having to discuss a survival plan based on ten items that could be rescued from the plane. The main task for the group consists of assigning priorities to each of these items, based on how useful they seem for helping the group to survive. This, however, is not trivial, since it involves creative and strategic thinking. A mirror can, for example, be used to signal for help or to make fire; a parachute can give shelter from the sun and can help to gather water at night; a compass may only be useful if the group decides to leave their current location in search for rescue.

This type of task is *ill-defined*, since the group is unlikely to have the required expertise or experience to rationally rank the items, and can only resolve their uncertainty by collaborating and jointly exploring the potential use and value of every individual item. According to media richness theory (see Chapter 2.2.1), the richer the medium, the better it supports uncertainty resolution. Therefore, it can be expected that this type of task is suitable to discriminate between media with different levels of richness, and is thus useful for this experiment.

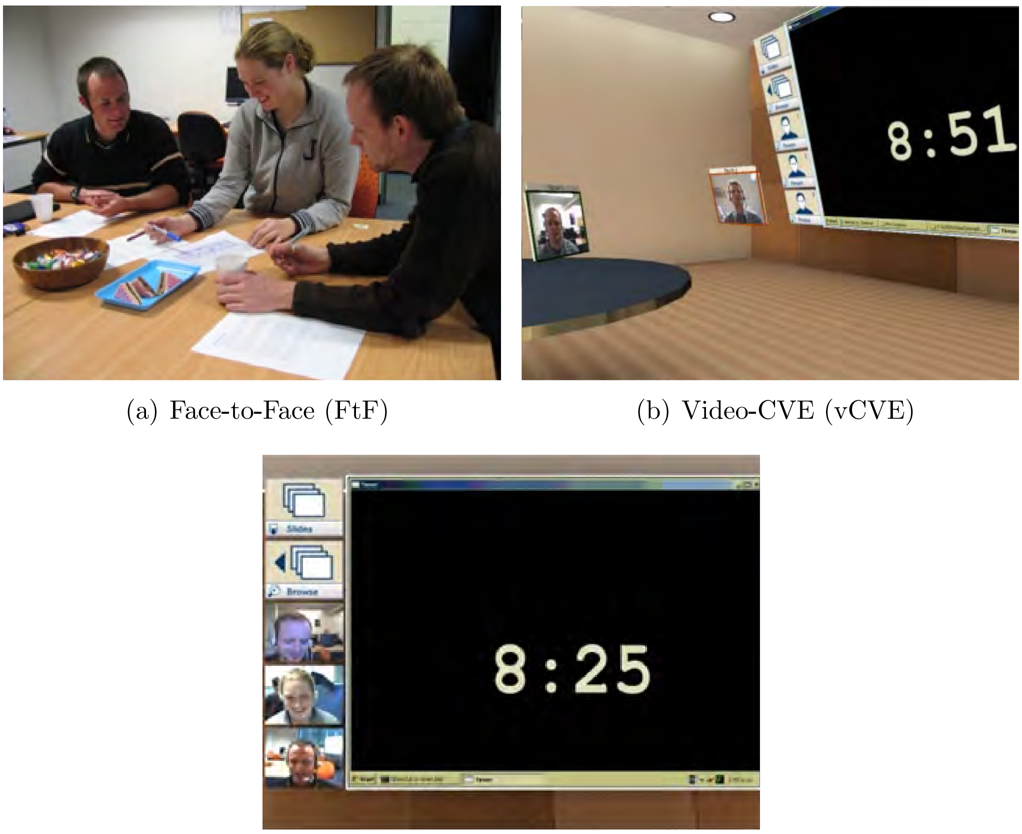
A detailed game description was obtained from a web resource for scouting games[1] and was adapted in the following ways: (1) the number of items on the ranking list was decreased to 10 (from 15 in the original task), (2) values in miles and Fahrenheit were converted to Kilometers and Centigrade.

[1] http://www.rogerknapp.com/download/games.htm. Last accessed: May 2005.

4.1.2 Experiment conditions

The experiment followed a within-subjects design, meaning that every group worked on the given task in each of the following three conditions.

In condition "Face-to-Face" (Figure 4.1(a)) participants were co-located in one room, sitting around a table. In condition "standard video-conferencing" (Figure 4.1(c)) as well as in condition "video-CVE" (Figure 4.1(b)) participants were located in separate rooms, from where they communicated with the others via one of the two video-conferencing setups.



(a) Face-to-Face (FtF)

(b) Video-CVE (vCVE)

(c) Standard video-conferencing (sVC)

Figure 4.1: Conditions in the experiment "Desert Survival Game".

Both mediated conditions (sVC and vCVE) were implemented as variants of "cAR/PE!", previously introduced in Section 3.1.1. Condition vCVE applied cAR/PE! as-is. This means that all participants were represented by their video-avatars which they could freely navigate ("head" rotation left/right and up/down, and translatory avatar movement forward/backward and left/right) through the 3D virtual room using a normal computer mouse. That way participants could place themselves virtually close or far from each other, could face each other or the projection wall of the room, or could for example choose to "sit" around a virtual table. If a participant got "lost" in the virtual room he or she could get back to a default position at the virtual table by clicking a home-button shown in an interactive menu on the bottom of the screen (not visible in Figure 4.1(b)). Integrated into the virtual environment was a virtual screen which displayed the remaining time for the current round.

In condition sVC the view into the environment was locked at a fixed position facing the timer screen (see Figure 4.1(c)). Video streams of all participants were displayed beside the timer screen, comparable to other conventional desktop video-conferencing layouts which follow the "what you see is what I see" (WYSIWIS) paradigm.

4.1.3 Apparatus

For the remote conditions, three acoustically and visually separated rooms were prepared with identical standard desktop PCs, monitors (TFT, 17", 1280x1024 pixel resolution), head-sets (stereo with mono microphone), and webcams (USB, CIF resolution) (see Figure 4.2). All three PCs were connected via a standard 1 Gb Ethernet network switch. In both remote conditions the same video and audio codecs were used. Video and audio were synchronised with a latency (loop) of about 300 ms. The size of the video for each participant in the sVC condition was 6 cm x 4 cm as measured on the monitor screen. In the vCVE condition this size varied according to the movement of the participants.

4.1.4 Participants

Forty-two subjects (36 male and 6 female) participated in the experiment. In 14 sessions, each of the three participants of a team took part in three trials, giving a total of 126 trials. The age of the participants ranged from 19 to 63 years (median age 33 years). They had no prior knowledge of the experiment except that the objective was to compare video-conferencing systems.

Participants were recruited by personal invitation mainly out of Information Science staff members. The assignment of participants to groups and time slots was pseudo randomised, as every participant could individually choose an available time slot in an assignment form. Two additional subjects had to be recruited "on-the-fly" as a substitute, because two participants did not arrive on time.

Each session lasted for approximately one hour, including introduction, three trials $\times$ 10 minutes, and the time to fill out the questionnaires.

4.1.5 Procedure

The experiments were conducted in May 2005 at Otago University in New Zealand. A group of three subjects participanted in each one-hour session. Upon arrival the participants could choose one of three seats at a table (marked as person 1, 2, 3) and were asked to read the Participant Information, explaining (1) the goal of the experiment (investigating differences of video-conferencing systems), (2) the general procedure, (3) the anonymity of the experiment, and (4) a participant consent form which was to be signed by the subjects. Additionally, the document contained a General Demographics Questionnaire. After completion, the participant instruction sheet was given out, which described the Desert Survival Game.

Each group participated in three rounds, one for each of the three conditions. The order of conditions was controlled using a Latin square. The task in each condition was the same (ranking of item list). To keep the participants interested and involved across three conditions, they were told the calculated difference between their ranking and an expert solution after each of the three conditions. This gave some feedback on how well their team was doing and encouraged them to further improve their result in the next round. Since this experiment primarily investigated the support of the collaboration process and not the outcome, the actual quality of a group's final ranking or any form of learning effects associated with the task were of little concern.

One participant in each condition had the role of the "scribe" who got the list of items and who had to write down the item rankings the group agreed on in each round. Each of the three participants took on that role in one of the three rounds. This was randomised by having participants who have not yet been the scribe roll a dice before each round. The participant with the highest number was chosen for the scribe role in the following round.

In the vCVE round, participants were given an introduction of how to use the mouse for navigation in the virtual room and got another two minutes to make themselves familiar with it. For further reference a sheet was put at the workplace which graphically explained the mouse interface.

The sVC condition did not require any instruction. In both mediated conditions the subjects wore audio head-sets (see Figure 4.2) which were explained and adjusted for best comfort.

After each condition the participants came together and filled out the experiment questionnaire (see Appendix A.1) on paper. The interim score of their ranking was announced and they continued to the next condition trying to further improve their result.

Figure 4.2: A video-conferencing workspace: the participant is communicating with the others through a web-camera and a headset.

The experimenters played a passive role. They were instructed only to interfere and help the participants in case of unforeseen circumstances or if any assistance with the technical equipment was needed. During each round, experimenters made notes of any salient events.

4.1.6 Measures

This experiment assessed differences between the experiment conditions based on the self-reported experience of the users. The experiment questionnaire (attached in Appendix A.1) included:

- Nine bipolar pairs following Short et al.'s (1976) social presence measurement approach (see Section 2.2.1).
- Thirty-eight items of Biocca et al.'s (2001) Networked Minds (NWM) measure of social presence (see Section 2.2.1).
- Eighteen exploratory items. These included four attitude test items that have been applied in video-conferencing studies by Champness (1973, as cited in Short et al. 1976), and fourteen items targeting subjective communication quality and task performance.

4.2 Data analysis

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

4.2.1 Social presence as quality of the medium

Short et al. (1976) conceive social presence as a uni-dimensional property of the communications medium. An effect for social presence following their definition was tested by comparing the social presence average score of the three conditions in an analysis of variance with "Medium" as a within-subject factor. If a significant main effect was found, conditions were also compared pairwise using the Bonferroni adjustment for multiple comparisons.

4.2.2 Social presence as the experience of the other

Biocca et al. (2001) conceive social presence as a multi-dimensional psychological state. An effect for social presence following their definition was tested by comparing the averages of every sub-factor in an analysis of variance with "Medium" as a within-subject factor. If a significant main effect was found, conditions were also compared pairwise using the Bonferroni adjustment for multiple comparisons.

4.2.3 Exploratory questionnaire items

The 18 exploratory items were first subjected to a *principle components factor analysis*. An effect for every obtained factor was then tested by comparing the factor average scores of the three conditions in an analysis of variance with "Medium" as a within-subject factor. If a significant main effect was found, conditions were also compared pairwise using the Bonferroni adjustment for multiple comparisons.

Since this method of principle component factor analysis is not commonly used computer science, it is explained here in more detail.

The goal of a principle components factor analysis is to describe a set of p variables X_1,X_2,...X_p in terms of a smaller number of indices or factors, and hence elucidate the relationship between these variables (Manly 1986, page 93). In the analysis of questionnaire data, for example, factor analyses are used for the purpose of data reduction to identify a small number of factors that explain most of the variance that is observed in the much larger number of correlating items. The principle components factor analyses performed in this as well as in other experiments in this thesis comprised three steps:

1. *Factor extraction:* principal components analysis is used to obtain a provisional factor solution. Factor components are extracted by formation of uncorrelated linear combinations of the observed variables. The provisional factor solution contains as many factor components as there are variables. The first component has maximum variance. Successive components explain progressively smaller portions of the variance. A decision about how many factors to retain for the final solution is

made by the analyst, based on the amount of variances explained by the strongest components. If a set of data contains highly correlating variables, a high percentage of the observed variance can be explained by only a few strong factors. Further factors can then only account for a small proportion of the variation in the data and may reasonably be ignored. A general rule of thumb is to only extract as many factors as there are eigenvalues greater than unity for the correlation matrix of the test scores, since a factor associated with an eigenvalue "explains" less variation in the overall data than one of the original test scores.

2. *Factor rotation:* the extracted factors are then rotated by an orthogonal "Varimax rotation", which minimises the number of variables that have high loadings on each factor. This allows for easier factor interpretation.

3. *Factor interpretation:* the final, rotated factors are interpreted and labelled based on the variable(s) with the highest factor loadings.

4.2.4 Scale validity

A principle component factor analysis was also performed with the data set obtained for Short et al. (1976) measure of social presence. This was done to test whether social presence would indeed get extracted as the only factor, or, like observed by Short et al. (1976) in some of their studies, if any other factors would emerge.

The reliability of both social presence scales were afterwards tested by calculating *Cronbach's alpha* for every unidimensional construct. Cronbach's alpha is a standard measure of the reliability of a psychometric instrument. The value for Cronbach's alpha generally increases when the correlations between the items increase and therefore indicates the level of internal consistency of a scale. As a rule of thumb, a Cronbach's alpha between 0.8 and 0.9 marks a sufficient reliability, alphas greater than 0.9 mark a high reliability (Bortz and Döring 2003, page 198).

4.2.5 Age and simulator experience effects

Average factor scores of both social presence measures were also subjected to a 2×2×2 mixed factor ANOVA with the within-subjects factor "Medium" and the between-subjects factors "Simulator experience" and "Age group". These tests were purely exploratory, mainly to investigate whether the simulator experience or the age of participants showed some unwanted effects or interactions, and thus need to be controlled during the next recruitment process.

4.3 Results

4.3.1 Social presence as a quality of the medium

The mean values and standard errors of social presence assessed through the semantic differential scales are displayed in Figure 4.3.

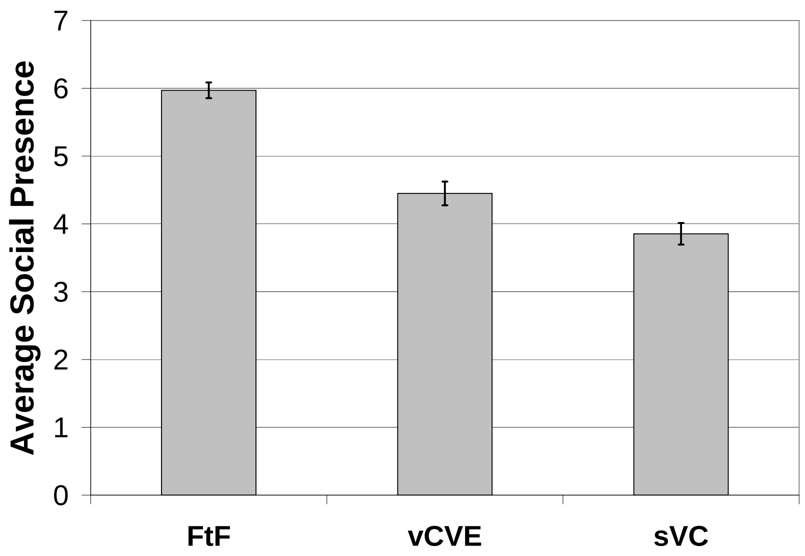


Figure 4.3: Average ratings and standard error for social presence based on the semantic differential technique.

There was a significant main effect of "Medium" , $F(2,82)=67.2$, $p<0.001$. Post-hoc pair-wise comparisons furthermore revealed that social presence of FtF (M=6.0, SD=0.75) was rated significantly higher than vCVE (M=4.5, SD=1.1, p<0.001) and sVC (M=3.9, SD=1.0, p<0.001). Social presence of vCVE was also rated significantly higher than social presence of sVC (p=0.005).

4.3.2 Social presence as the experience of the other

The calculated factor scores and standard error for every sub-factor in the NWM measure of social presence are shown in Figure 4.4. The sub-factors "Isolation/Inclusion" and "Dependent Action" were not included in the analysis, since they failed the test for internal consistency (see Section 4.3.4).

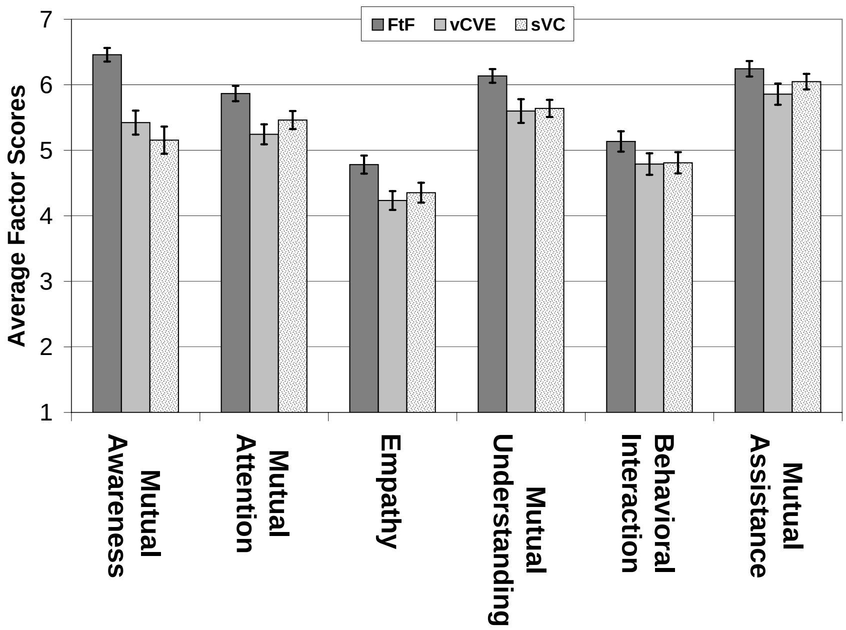


Figure 4.4: Average ratings and standard errors of each factor of the NWM measure of social presence.

A significant main effect of "Medium" was found for the factors Mutual Awareness, Mutual Attention, Empathy, and Mutual Understanding (see Table 4.1). Pairwise post-hoc comparisons revealed a consistent pattern of significant differences between FtF and vCVE and between FtF and sVC, but not between vCVE and sVC in all factors.

Factor	**F**	**df**	**p**	**Post-hoc comparisons**
Mut. Awareness	22.8	2	<0.001	FtF>vCVE; FtF>sVC
Mut. Attention	8.2	2	0.001	FtF>vCVE; FtF>sVC
Empathy	7.0	2	0.002	FtF>vCVE; FtF>sVC
Mut. Underst.	7.2	2	0.001	FtF>vCVE; FtF>sVC
Beh. Interaction	2.8	2	not sig.	–
Mut. Assistance	3.0	2	not sig.	–

Table 4.1: Results of the analyses of variance and post-hoc tests of social presence sub factors.

4.3.3 Exploratory questionnaire items

The principle components factor analysis extracted five factors which accounted for 72% of the variance (see Appendix B.1 for details).

The factors were labelled according to the item with the highest factor loading. Table 4.2 contains the names of the extracted factors, the number of items which belong to a factor, its internal consistency, and examples.

Factor (Items)	α	**Example**
Realism (4)	0.92	It was like a face-to-face meeting
Ease of Communicat. (6)	0.86	I felt like I was often interrupted
Turn Flow (2)	0.85	I waited for my turn to speak
Speaker Pauses (1)	–	There was a lot of time when no one spoke
Media Impact (1)	–	The task performance was affected by the medium

Table 4.2: Results of the factor analysis of 18 exploratory questionnaire items: names of extracted factors, number of items, their internal consistency, and examples.

Figure 4.5 shows the average factor scores and standard error for every extracted factor.

A significant main effect of “Medium” was found for the factors “Realism”, “Ease of Communication”, and “Turn Flow” (see Table 4.3).

Pairwise comparisons revealed significant differences between all conditions in the factor “Realism”. FtF felt the most like a real meeting, followed by condition vCVE, which in turn felt more like a real meeting than condition sVC.

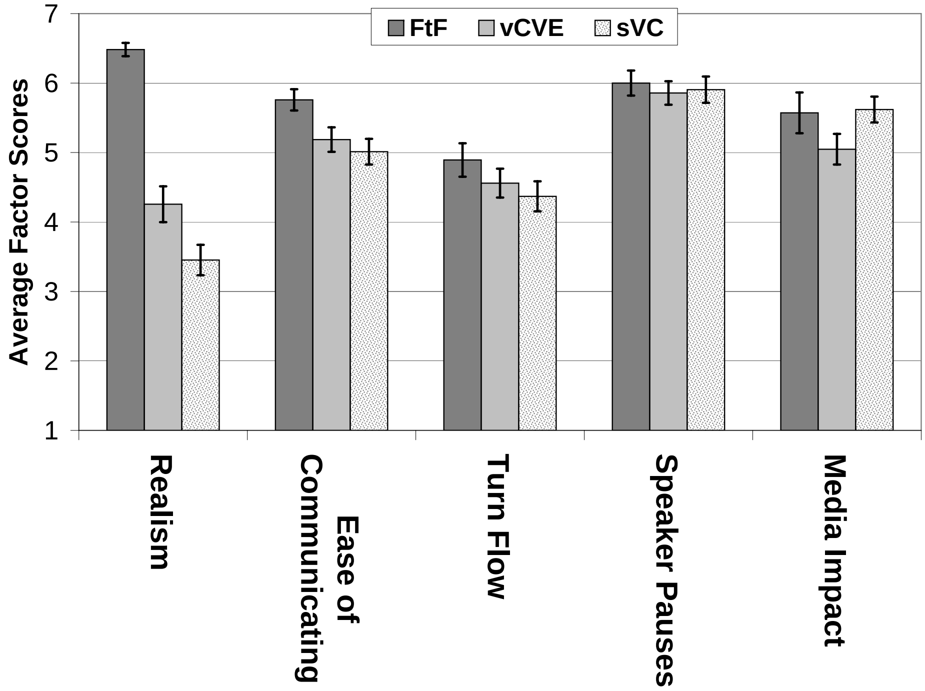


Figure 4.5: Average and standard errors of extracted factors.

Pairwise comparisons also revealed that the "Ease of communication" was rated higher in FtF than in both video-conferencing conditions.

Factor	**F**	**df**	**p**	**Post-hoc comparisons**
Realism	67.6	2	<0.001	FtF>vCVE>sVC
Ease of Com.	9.8	2	0.000	FtF>vCVE; FtF>sVC
Turn Flow	3.2	2	0.044	
Speaker Pauses	0.18	2	not sig.	–
Media Impact	2.0	2	not sig.	–

Table 4.3: Results of the analysis of variance and post-hoc tests.

4.3.4 Scale validity

The semantic differential scale

The social presence data gathered with the semantic differential technique were subjected to a principle components factor analysis. This analysis extracted one single factor, social presence, which accounted for 67% of the variance, confirming the assumed uni-dimensionality of the construct. Table 4.4 contains the nine scale items and their corresponding factor loadings, that is, the correlations between every individual item and the social presence factor.

Bipolar Pair	**Factor Loading**
cold–warm	0.87
insensitive–sensitive	0.90
colourful–colourless	0.68
impersonal–personal	0.84
unsociable–sociable	0.87
small–large	0.77
ugly–beautiful	0.75
open–closed	0.85
passive–active	0.81

Table 4.4: Factor loadings for social presence, measured with the semantic differential technique.

The item “colourful – colourless” had a substantially lower factor loading than the other items, possibly because participants associated it with graphical features of the interfaces rather than with the inherent nature of the communication they support. It was therefore dropped from further analysis.

Cronbach’s alpha of the social presence scale based on the remaining eight items was 0.93, which suggests high reliability.

The NWM measure of social presence

Cronbach’s alpha was calculated for every individual sub-factor of the NWM measure of social presence separately. Data from all three conditions were considered. Table 4.5 shows the obtained alpha values of this study as well as the one reported in Biocca et al.’s (2001) pilot study for comparison.

The factors “Isolation/Inclusion” and “Dependent Action” show insufficient internal consistency and were therefore excluded from further analysis. The remaining factors showed sufficient reliability. These results replicate the factor validity of the pilot study found by Biocca et al. (2001).

Factor	Alpha	Alpha Biocca et al. (2001)
Isolation/Inclusion	0.58	Insf.
Mutual Awareness	0.84	0.81
Mutual Attention	0.76	0.82
Empathy	0.70	0.76
Mutual Understanding	0.86	0.87
Behavioural Interaction	0.83	0.75
Mutual Assistance	0.73	0.69
Dependent Action	0.40	Insf.

Table 4.5: Factor reliability of the NWM measure of social presence in this and Biocca et al.'s (2001) pilot study.

4.3.5 Age and simulator experience effects

The social presence factor scores of both social presence measures were also explored for unwanted interactions and between-subjects effects with the control factors "Simulator experience" and "Age group". Simulator experience included any experience with 3D computer games.

An exploratory 3×2×2 mixed factor analysis with factors Medium, Age group and Simulator experience was performed. The factor "Age group" had the two levels "younger half" and "older half", the factor "Simulator experience" had the two levels "no experience" and "some experience". Participants were categorised based on the data they provided in the General Demographics Questionnaire at the beginning of the experiment.

No significant interactions or between-subject effects were found.
This result was based on 22 participants in the older age group and 20 participants in the younger age group. Twenty-four of the participants indicated that they had no previous simulator experience, while 18 participants did have some experience.

Although it would have been of interest to also test for effects of gender and video-conferencing experience, these factors were not considered, since the population sizes for their levels were too unbalanced.

4.3.6 User observation in condition vCVE

This section lists some salient observations that were noted by experimenters.

Navigation: Turning one's head in reality is fast, natural, and normally does not involve a lot of mental or motor skills. Turning the head of one's video-avatar in the vCVE condition, however, caused some problems. In many cases, the mouse interface provided was not easy or fast enough.

Consequently, some users changed their avatar's position and orientation in the virtual room only rarely. Often, to avoid the need for virtual head-turning, they tried to navigate their avatars into a "comfortable view" position from where they could see the two other video-avatars at the same time. This led to the repeatedly occurring navigation sequence as depicted in Figure 4.6. It shows the schematic representation of the avatars P1, P2, and P3 including their view frusta.

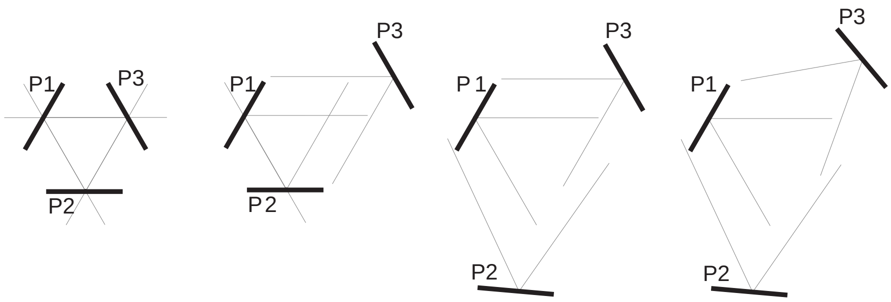


Figure 4.6: A frequently observed navigation sequence. Avatars P3 and P2 navigate to positions from where they can see both other avatars at the same time, leaving P1 with the need for more head turning to see either one or the other avatar.

Initially, as depicted at the left side of the figure, all three avatars are arranged in a circle around a virtual table. In that position, every participant can see one half of the other video avatars. Then, to get to a position where he or she can see both of the other avatars in full, P3 in the given example moves backwards. This, however, causes the avatar P3 to disappear from the view field of both P1 and P2. Therefore, also P2 (or P1) moves backwards looking for P3 and trying to find a position where both P1 and P3 are visible at the same time. This, in turn, causes P2 to partly disappear from P3's view, which is corrected by P3 with a further small push back combined with a small turn to the left. The resulting arrangement is shown on the right of the figure. In this configuration, both P2 and P3 assume "comfortable positions", leaving P1 with the burden of having to actively turn his or her avatar to see either P2 or P3 at a time. If P1 tries to find a comfortable view and moves backwards a configuration similar to the one on the left of the figure is reached and the sequence starts in a similar way again.

The underlying problem was that participants often assumed full view reciprocity following a "if I can see you, you can see me" paradigm while navigating their avatars. This, however, is not the case in a CVE due to a restricted field of view compared to the real world. Consequently, participants did not realise that they moved out of the view of the others when adjusting their own views.

Other forms of navigation have been proposed by participants who frequently play computer games. They suggested they would have preferred a keyboard or combined keyboard/mouse interface for navigation since that would be more "natural" to them.

Immersion: There were clear indicators that users understood the spatial character of the interface in the vCVE condition. For example, users frequently turned their avatar away from the other avatars towards the projection screen to see the timer and then back again, similar to a glimpse at the clock in a real room. Users also clearly liked and exploited the spatial sound, for example by adjusting the view direction towards an avatar that was temporarily out of the view while talking to them.

4.4 Discussion

The obtained significant difference between the two video-conferencing interfaces for social presence measured following Short et al.'s (1976) approach is an encouraging result from two perspectives. Firstly, it suggests that the semantic differential approach to measuring the social presence of a medium is sensitive enough to discriminate between different video-conferencing interfaces. This makes it a suitable metric for the research focus of this thesis. Secondly, the fact that social presence of the video-CVE was rated higher than standard video-conferencing aligns with the research hypothesis of this work and provides a first indication that collaboration in video-CVEs feels more like being face-to-face than standard video-collaboration.

The data collected with the NWM measure of social presence could reproduce most of the results reported in Biocca et al.'s (2001) pilot study. However, the measure could not identify any differences between the two VMC interfaces. This result indicates that measures of the experience of social presence which are entirely based on self-reported introspection may only be sensitive to large effects that occur between substantially different media. Since this thesis intends to investigate more subtle effects between variants of VMC systems, this measure therefore is not suitable.

The results for the exploratory items targeting subjective communication quality and task performance proved to be a poor way to assess these dimensions. This suggests that they should therefore be replaced by more sensitive process and product measures in following experiments.

The only stable factor that emerged which showed an effect also between the mediated conditions was "Realism". This factor assessed the perceived resemblance of a mediated situation to a real face-to-face meeting and produced meaningful results for the conditions of this experiment. The underlying four items were taken from a scale of Champness (1973, as cited in Short et al. 1976) who evaluated a commercial video system. The relevance of attitudinal items such as these for social presence studies has been pointed out by de Greef and IJsselsteijn (2001). The results obtained confirmed this and suggested to include similar scales in further studies.

The exploration of demographic factors showed not salient interactions. This suggested that age and simulator experience do not have to be controlled or considered as extra factors in following experiments.

4.5 Experiment summary

The experiment "Desert Survival Game" investigated the utility of several subjective rating scales for comparisons between standard video-conferencing systems and video-CVEs. It produced a range of useful results.

Two social presence scales which assess facets of the user experience in mediated communication were tested for their applicability in comparative studies involving a standard video-conferencing system, a video-CVE, and a face-to-face condition. The first measure applied selected bipolar pairs as proposed by Short et al. (1976). This approach proved to be a reliable and sensitive metric to assess social presence, which in that case is conceived as a subjective quality of a medium. Participants assigned a higher level of social presence to the video-CVE than to the standard video-conferencing interface. The second measure applied the Networked Minds questionnaire proposed by Biocca et al. (2001), which determines social presence as a multi-factorial, psycho-phenomenal state. The data collected with this questionnaire reproduced most of the results reported in Biocca et al.'s (2001) pilot study including a general capability to discriminate between a standard video-conferencing condition and an unmediated face-to-face situation. However, the measure could not identify any differences between both mediated conditions.

An exploratory questionnaire including 18 items also assessed the subjective communication quality and task performance. However, no useful results other than the emergence of a "Realism" factor could be obtained from this analysis.

The results obtained using the NWM measure of social presence and the perceived communication quality and task performance showed that the instruments were not mature enough, and therefore may only be able to discriminate between two media that are substantially different from each other. For the more subtle differences between different interfaces which are of primary interest of this research work, however, they proved not to be applicable and should therefore be replaced by other, preferably objective, process and product measures.

Observations of participants furthermore revealed some usability issues in the video-CVE condition. First, participants did not like the mouse-based navigation interface, which inhibited them from navigating their avatars to the extent that was anticipated. Second, the restricted field of view of an avatar led to problems in the group arrangement of the avatars, since there was no constellation in which every participant could see both of the other participants' avatars at the same time.

Taken together, these initial findings were encouraging for two reasons. First, a measure was identified that suited the focus of this research. And second, the differences that emerged characterised collaboration in video-CVEs to be more similar to being face-to-face, which aligns with the general hypothesis of this research.

The next chapter will describe a follow-up experiment which re-applied Short et al.'s (1976) social presence measure in a different collaborative task to test its reliability and confirm the main finding of this experiment. Furthermore, it applies a physical presence measure to investigate the impact of the level of transportation of a medium.

Chapter 5

Experiment: "Dream House"

The previous chapter reported details of the first experiment exploring the utility of several subjective rating scales for the focus of this research. The main finding was that one of two social presence scales could successfully discriminate between a standard video-conferencing system and a video-CVE in a conversational situation, which suggested its applicability in future comparisons involving different VMC interfaces.

The experiment "Dream House", presented in this chapter, was designed to investigate the repeatability of this result in a comparative study involving the same type of interfaces, but using a different type of collaborative task. Based on the initial result, the main hypothesis to be tested was therefore more specific: that social presence of video-CVEs is perceived to be higher than social presence of a standard video-conferencing interface, and thus that collaboration in video-CVEs feels more like being face-to-face.

In addition, this experiment also explored the question if the perceived transportation aspect of a video-CVE is beneficial for promoting social presence. The level of transportation of a communication medium is determined by the degree to which an interface "transports" the presence of a user to a remote space in order to come together with others (see Section 3.2.2). This question was investigated in two steps. First, the degree to which participants actually felt transported to a remote space as manifested in a sense of physical presence (often paraphrased as a feeling of "being there") at that space (see Section 2.4.1) was measured. Second, a correlation between the sense of physical presence and social presence was explored.

Finally, participants' comments and preference rankings were assessed in informal interviews to gather complementary data that could help interpret the other findings.

5.1 *Experiment design*

This experiment was again a joint effort of the author and members of the Multi-Media Systems Research Laboratory (MSRL) at the University of Otago. The following sections only present the objectives and results that were of interest to this research. Further results can be referred to in Hills' (2005) Honours thesis.

The experiment followed a within-subjects design with two conditions. Participants worked on a given task in groups of three, where one participant was an experimenter who took on the more passive role of a moderator.

5.1.1 Experimental task

The task was of judgemental nature. Participants had to evaluate five different house designs presented to them and then agree upon the best candidate according to a given problem definition.

To make the review of these design alternatives more realistic and comprehensible for the participants, it was "staged" as the following situation: participants were asked to imagine that they had just won lotto and therefore wanted to buy their "dream house". They were also told they were in a family of six (Grandma, Mum, Dad, Teenager, and themselves) who will all live in that house together. The reason they met on that day was to discuss advantages and disadvantages of a selection of five house designs that they were to consider for purchase, and to ultimately pick their favourite candidate.

The different house alternatives were created and made available by design students in the course of a design competition held at the Technical University Eindhoven in the Netherlands. Every house design represented a different European country.

The task was split into two parts so it could be worked on in both conditions. In the first part participants were advised to get a broad overview of all the models and to narrow the list of potential designs down to three. In the second part they were then asked to re-evaluate their favourites in depth and make their final decision.

Participants could take up to eight minutes in each round. However, they could finish the round before in case they came to an early agreement.

5.1.2 Experiment conditions

Two interfaces were implemented as variants of the cAR/PE! virtual tele-collaboration space (see Section 3.1.1).

Figure 5.1: The condition "video-CVE" (vCVE) in the experiment "Dream House". Two participants inspect and discuss the architectual design of a virtual house.

In the condition "video-CVE"(vCVE), shown in Figure 5.1, participants could control their individual view into the virtual scene by navigating their video-avatars using a computer mouse. The moderator put the models of the houses which were to be discussed at the centre of the table, where they could then be inspected by the two participants from all sides. While inspecting, participants could point out features of the model to the other participant with a telepointer, which appeared when they moved their mouse cursor over the model.

The interface provided several navigation mechanisms which suited the review. Participants could for example easily "spin" around the table and the house models by using a special feature of the navigation interface. Furthermore, they could select predefined perspectives (e.g. bird's eye overview of the room or default position at the table) into the virtual room by clicking on navigation shortcut buttons (not visible in the figure).

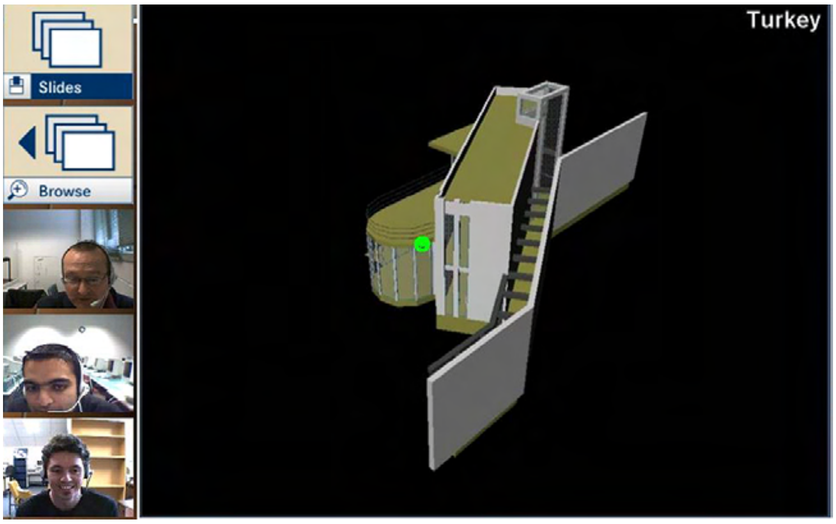


Figure 5.2: The condition "standard video-conferencing" (sVC) in the experiment "Dream House". Two participants and the moderator review and discuss a house design displayed on the shared presentation wall.

The interface in the condition "standard video-conferencing" (sVC), shown in Figure 5.2, locked the view of all participants so that they were all looking toward an interactive wall. There, the participants' videos were displayed next to a shared slide show which contained slides of four different perspective shots of each of the houses. The participants could flick through these slides either by pushing the "PageUp/PageDown" buttons on their keyboard, or by clicking a menu button next to the presentation wall. Like in the other condition, participants could use a telepointer to point at features of the individual house designs as shown on the slides. At all times, the participants could be sure that all other participants would see the exact same view.

5.1.3 Apparatus

Three acoustically and visually separated rooms were prepared with identical desktop PCs and monitors (TFT, 17", 1280x1024 pixel resolution). Further equipment at each work station included a USB webcam with CIF resolution and a regular stereo head-set with an integrated mono microphone. A standard 1Gb/sec network switch connected the computers. Figure 5.3 shows a participant at one of the work stations.

Figure 5.3: A participant sitting at one of the workstations, discussing the house designs with her team partner, who was located at a similar workstation in another room.

5.1.4 Participants

Thirty-six volunteers (26 male and 10 female) were recruited mainly from staff and students at the University of Otago for this experiment. Eighteen sessions were conducted where two subjects participated in two trials which gave a total of 72 trials. Participants were between 17 and 50 years old (median age 26) and were naive to the research focus of this experiment and only knew that the experiment's objective was to compare video-conferencing systems.

The participants of eight of the eighteen groups considered themselves to be friends, whereas participants in the other 10 groups stated that they had never met before or had only passing knowledge of each other.

5.1.5 *Procedure*

At the beginning of the experiment, the participants were first introduced to each other in case they had not met before, were then asked to sign a consent form, and to fill out a General Demographics Questionnaire. After that, two experimenters lead them to the work stations and instructed them on how to use the controls of the video-CVE in a scenario which had nothing to do with the experiment task. About ten minutes later, they came together again and received the task description.

After clarifying any concerns about the task, they were sent back to their work stations from where they started to review the houses in the first condition. The order of the conditions alternated between all groups to prevent ordering effects. Once the participants signalled they had completed the first part of the task, they came back together and filled out the prepared questionnaire.

They then continued with the second part of the task in the other condition which had been set up for them during the time they filled out the questionnaires. Once the team finally agreed which of the houses would be their "dream house", the second round was over and the participants were asked to fill out the same questionnaire again.

In condition cVE, the moderator placed the models on the table for the participants, so that they could focus on their conversation and navigation. During both conditions, the moderator also provided some "sparks" to keep conversations between the participants going. This was done for example by asking the participants questions like "Do you think grandma would like this house?", or "Which rooms would you give to the kids?".

Before the participants were dismissed, they were asked which of the conditions they preferred and were briefly interviewed for further comments on the interfaces.

The total duration of each session was between thirty and forty minutes.

5.1.6 *Measures*

The measures used to assess social presence and physical presence are explained in the following two sections.

Social presence

Social presence was measured with ten bipolar pairs (see Appendix A.2, items 13 – 22) following Short et al.'s (1976) measurement approach.

The previous experiment found this measure to be sensitive enough to discriminate between a video-CVE and standard video-conferencing interface in a conversational task. However, the item *colourful – colourless* showed a lower factor loading than the other items, possibly because participants associated it with graphical features of the interfaces rather than with the inherent nature of the medium.

The questionnaire in this study therefore replaced the items *colourful–colourless* and *ugly – beautiful*, which are both of a graphical nature, with the two items *constricted – spacious* and *dehumanising – humanising*, which have been applied in other studies by Short et al. (1976). Another item, *boring – interesting* was added which was devised by the author.

Physical presence

The sense of physical presence (see Section 2.4.1) was assessed by the Igroup Presence Questionnaire (IPQ) (Schubert et al. 2001). Out of the thirteen items of the original scale, twelve were included in the experiment questionnaire in the form of five-point Likert scales (see Appendix A.2, items 1–12). One item was dropped because it explicitly asks for the experience of free navigation through the virtual environment, which was not provided in the condition sVC.

5.2 Data analysis

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

5.2.1 Social presence

Average factor scores for social presence were calculated and tested for significant effects between conditions vCVE and sVC in a paired-samples t-test (2-tailed).

5.2.2 *Physical presence*

Average factor scores were also calculated for physical presence and tested for significant effects between conditions vCVE and sVC in a paired-samples t-test (2-tailed).

5.2.3 *The relation between physical presence and social presence*

The relationship between social presence and physical presence was explored by inspection of a scatter plot which establishes a relationship between the factor average differences of social presence and physical presence across both conditions. By plotting the difference scores, basic relationships between the tendencies of the two types of presence emerge and can be analysed optically. This method for comparing social presence to physical presence has, to the author's knowledge, not been used in this form in similar studies before and is therefore unique.

5.2.4 *Reliability analysis*

As part of the reliability analysis for the social presence scale, all bi-polar pairs (including three new ones) were tested for their factor loadings on social presence. The collected social presence data was therefore subjected to an exploratory factor analysis. The method of extraction was a principle component analysis with Varimax rotation. In addition, Cronbach's alpha as a measure of internal consistency was calculated for social presence and physical presence.

5.2.5 *Preference ratings*

The participants' preferences for one or the other interface and comments were assessed during a brief, unstructured interview with the participants at the end of the experiment.

5.2.6 *Exploration of other factors*

Average factor scores of social presence and physical presence were also subjected to a 2×2×2 mixed factor ANOVA with the within-subjects factor

"Interface" and the between-subjects factors "Gender" and "Relationship".

These tests were purely exploratory, mainly to investigate whether the gender or relationship of participants showed some unwanted effects or interactions, and would thus need to be controlled during the next recruitment process.

5.3 Results

Thirty-five data sets from 18 sessions were analysed. Due to a misunderstanding between experimenters, one participant was recruited twice. The second data set of that participant was therefore dropped and is not included in the analysis.

5.3.1 Differences in social presence

There was a significant main effect. Participants rated the social presence of condition vCVE (M=5.4, SD=1.0) significantly higher than social presence of condition sVC (M=4.6, SD=1.2), $t(34)$= -4.1, $p<0.001$.

5.3.2 Differences in physical presence

There was a significant main effect. Participants experienced a significantly higher sense of physical presence in the condition vCVE (M=3.7, SD=0.7) than they did in condition sVC (M=3.2, SD=0.6), $t(34)$= 4.0, $p<0.001$.

5.3.3 Exploring the relationship between social presence and physical presence

The scatter plot displayed in Figure 5.4 establishes a relation between the differences measured in social presence and physical presence across both conditions.

The x-axis represents the differences of physical presence, calculated by subtracting the physical presence score of condition sVC from the physical presence score obtained from condition vCVE for every participant. For example, if a participant rated the sense of physical presence in condition sVC on average as 3.8, and the sense of physical presence in condition vCVE

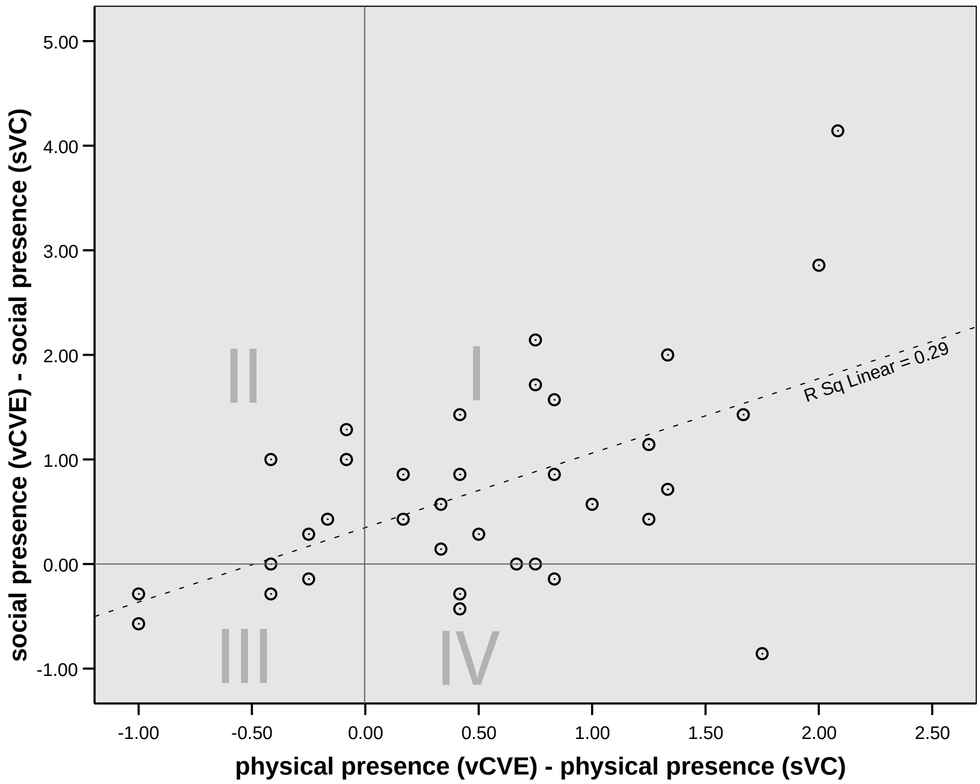


Figure 5.4: The relation of physical and social presence difference scores.

as 4.3, the difference score would be 0.5. In the same manner, the y-axis represents the difference score obtained by subtracting the social presence score of condition sVC from that of condition vCVE.

Every circle in the chart represents a data point of one participant. As can be seen by the cloud distribution of all circles, there is no strong mathematical correlation between these dimensions (which was not expected either). However, by looking at the population of data points in the different quadrants, labelled I – IV, one can see for how many participants the increase/decrease of one type of presence involved the increase/decrease of the other.

For nineteen participants (54%) both physical and social presence increased in the vCVE condition (quadrant I). Five participants (14%) felt less physically present in the vCVE condition, but rated social presence higher in vCVE (quadrant II). Four participants (11%) experienced less physical and social presence in condition vCVE (quadrant III). Four participants (11%) had a stronger sense of physical presence in condition vCVE, while considering vCVE to support less social presence (quadrant IV). For three participants, social presence did not change independent of an increasing or decreasing sense of physical presence.

Taken together, the tendency of physical presence matched the tendency of social presence in 65% (data points in sector I and III) of all cases.

One factor which may have had an impact on the relation between physical and social presence was the relationship of the team members. Figure 5.5 shows the same scatter plot as before with the additional indication if a data point represents a member of a team of "friends" or a team of "strangers". What stands out is a linear correlation between social and physical presence for friends (linear regression analysis, R^2=0.80, df=14, p<0.001), where the tendency of social presence and physical presence match in 80% of all cases. The data points for strangers, however, are scattered. Quadrant IV is populated only by strangers.

These patterns suggest that a correlation between physical presence and social presence is more distinct for people that know each other well.

5.3.4 Reliability analysis

Social presence

The factor analysis of the social presence data extracted two factors, which accounted for 71% of the variance. Table 5.1 reports the factor loadings for each variable on factor 1 and 2. Each number represents the partial correlation between the item and the rotated factor. Items 1, 2, 4, 7, and 10 load highly on factor 1 (indicated by the bold face number). These items include the four items warm-cold, personal-impersonal, sensitive-insensitive, and sociable-unsociable, which were associated with social presence by Short et al. (1976). Factor 1 can therefore be clearly identified as "social presence".

Factor 2 is best described by items 3, 5, and 8 (indicated by the bold face number). With the item *constricted–spacious* having the strongest loading,

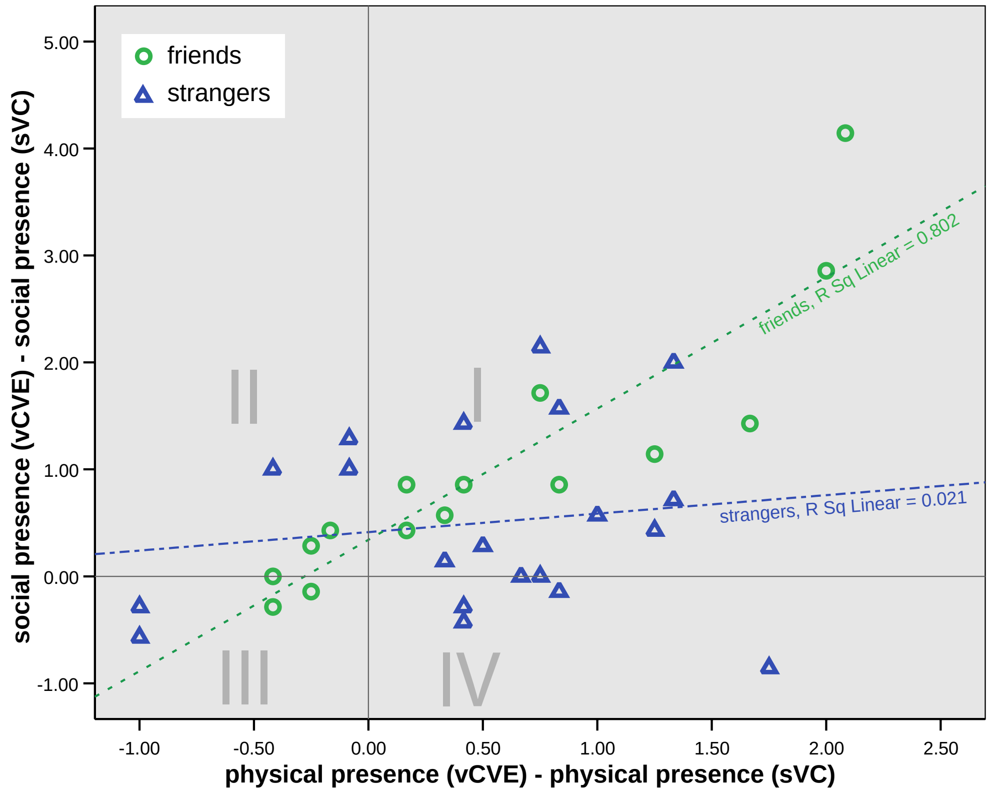


Figure 5.5: Differences in physical and social presence by relationship.

this factor bears some spatial character which might have resulted out of a direct association with the spatial features of the interfaces rather than with the inherent qualities of the communication medium.

For the analysis of "social presence" reported previously, the items that loaded on the factor 2 were not considered. Although items 6 and 9 loaded on both factors, they were considered in the analysis of social presence, since they loaded higher on factor 1.

A Cronbach's alpha of 0.90 was calculated for social presence being described by the remaining seven items, which indicated good scale reliability.

	Item	Factor 1	Factor 2
1	cold - warm	**0.70**	0.44
2	insensitive-sensitive	**0.77**	
3	small - large	0.32	**0.86**
4	impersonal - personal	**0.76**	0.30
5	closed - open	0.36	**0.82**
6	passive - active	0.60	0.47
7	unsociable - sociable	**0.84**	
8	constricted - spacious		**0.89**
9	boring - interesting	0.63	0.53
10	dehumanising - humanising	**0.77**	

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalisation.

Table 5.1: Rotated Component Matrix. Two factors were extracted. High factor loadings of items are emphasised.

Physical presence

To test the physical presence items for internal consistence, a Cronbach's alpha of 0.85 was determined which matches the reliability reported for this instrument in literature[1]. This value confirms a sufficient internal consistency.

5.3.5 User comments and preference

After the experiment, participants were asked which of the two conditions they preferred and were briefly interviewed about their general experience with each of the systems. Unfortunately, this data is only available for 28 out of the 35 participants due to some timing issues during four out of sixteen sessions (the next team showed up before the current team had finished, so no time remained for interviewing the current team).

Twenty out of the 28 participants interviewed said they preferred the condition vCVE for working on the given task. Six participants favoured the condition sVC. Two participant wanted a combination of both interfaces.

Those who preferred the condition vCVE commonly stated that they

[1] http://www.presence-research.org/, last accessed October 2007.

liked the "feeling" of being in a room and that they considered the features and interactivity based on a room as useful. They furthermore liked the way in which they could engage with the virtual objects, for example that they could "step inside" them. Some of them, however, also mentioned they had problems with the navigation.

The other participants saw the advantages of condition sVC in being less distracting and in providing a view where everyone can always see the same things. Furthermore, some thought it involved less effort to get an idea of a houses based on the prepared perspective shots displayed on the projection screen.

5.3.6 Gender and relationship effects

The interactions Interface×Gender and Interface×Relationship did not show significant effects. Nor was there a significant main-effect between different genders. However, there were main effects for the between-subjects factor "Relationship":

- Friends rated the social presence of both interfaces higher than participants that did not know each other well before the experiment $F(1,32)=10.3$, $p=0.003$.

- Friends furthermore perceived a higher level of physical presence in both interfaces than did participants that did not know each other well before the experiment $F(1,32)=16$, $p<0.001$.

This analysis was based on 10 female and 25 male participants, 15 participants considered themselves to be friends, whereas 20 participants had never met before or had only passing knowledge of each other.

5.4 Discussion

The hypothesis of this experiment was supported, namely the video-CVEs yield higher social presence than standard video-conferencing interfaces. This confirmed the result of the first study and strengthened the confidence into the utility and reliability of Short et al.'s (1976) measure of social presence.

The exploration of the correlation between social and physical presence produced a first weak indicator that a stronger sense of physical presence as induced by the "remoteness" of a medium may promote an increase of social presence associated with that medium. Taken together, the tendency of physical presence could predict the tendency of social presence in 65% of all cases. This prediction, however, was more accurate and increased to 80% if only data from participants who knew each other well before the experiment and who probably were familiar and comfortable with "being around" each other in real life as well. In contrast, the process of getting to know a stranger in a teleconferencing situation introduces some additional factors (e.g. a lower level of self-disclosure, lower level of common ground, higher threshold for the "suspension of disbelief") which may have impacted on the perception of non-verbal behaviour of a medium. The higher observed levels of social presence and physical presence for friends could be a further indication of this. As a consequence, participants in the next studies were recruited as teams of friends only to eliminate this heterogeneity of the population.

The factor loadings of the items applied in the social presence measure in this experiment clearly identified one factor as "social presence", but also extracted a second factor which may have picked up some spatial interface features rather than the interface's inherent qualities as a communication medium. These bipolar pairs should therefore only be applied with caution in experiments that investigate social presence in spatial communication media.

5.5 Experiment summary

This experiment investigated the difference of perceived social presence and physical presence between a standard video-conferencing interface and a video-CVE. Social presence was higher in video-CVEs, suggesting that collaborating in video-CVEs feels more like being face-to-face. This finding successfully replicated and strengthened the main result of the previous study.

Furthermore, the differences of social presence between interfaces were investigated for correlations with physical presence scores. In 65% of all cases, the tendeny of an increase or decrease of physical presence matched with the increase or decrease of social presence. This is but a small indication for a positive relation between the two types of presence. However, The correlation was much more distinct for participants who considered their team partner their "friend" in contrast to other teams made up of people that did not know each other well.

The majority of participants interviewed preferred the video-CVE over the standard video-conferencing interface in the given scenario.

The next two experiments, presented in the next two chapters, will extend the findings of this and the previous experiment in two ways. First, they will also consider objective measures to investigate the question, in which mediated condition participants not only *feel* but also *behave* like being face-to-face. Second, they will narrow the focus of the interface properties of video-CVEs and isolate the impacts of spatiality and transportation.

Chapter 6

Experiment: “Dogs and Owners”

This thesis investigates whether people who use video-CVEs for remote collaboration feel and behave more similarly to being face-to-face than if they use standard video-conferencing systems. The last two experiments revealed that video-CVEs support more social presence than standard video-conferencing interfaces. This is a first indication that video-CVEs push the user experience closer to face-to-face.

Video-CVEs differ from standard video-conferencing interfaces in their level of spatiality and transportation (see Section 3.2.2). The user experience (social presence and physical presence) was affected by these differences. However, observed effects could not be attributed to either spatiality or transportation, because the video-CVE conditions used in the first two experiments were both more “spatial” and more “transporting” than the standard video-conferencing interfaces.

The experiments presented in this and the next chapter disambiguate the causality of observed effects by concentrating on first the impact of adding spatiality to a VMC interface, and second on the impact of increasing the level of transportation of a video-CVE. The studies also considered objective measures that investigated collaborative behaviour.

The experiment “Dogs & Owners”, presented in this chapter, concentrated on spatiality in VMC. It looked at a more specific version of the main hypothesis of this thesis: that *adding spatiality* in VMC allows people to feel and behave in a way that is more similar to being face-to-face. Spatiality concerns an interface’s level of support for fundamental physical spatial properties such as containment, distance, orientation, and movement (see Section 3.2.2). The impact of spatiality of an interface on the user experience and user behaviour was isolated by the selection of a suitable scenario and the inclusion of a further spatial control condition.

The scenario for the experiment was chosen based on a common face-to-face collaborative situation, where aspects of spatiality play an important role: people who are gathered around a table to discuss a set of photos. The selection of this scenario was motivated by an ethnographic study conducted by Crabtree et al. (2004), who explored the embodied interactional ways in which people naturally collaborate around shared collections of photographs. They stress the importance for computer based photo-ware to support movement and manipulation of photographs, fine-grained gestures, and a range of awareness mechanisms that enable users to control access to and track the use of photographs. Table-top systems naturally support these requirements by the spatial properties given in any horizontal surface. People gathered around a shared tabletop surface naturally use these spatial properties to establish common ground and coordinate their activities (Scott et al. 2003, Kruger et al. 2004).

The conditions included two approaches for adding spatial cues to video-conferencing, a standard video-conferencing interface, and a face-to-face "gold-standard" control. One spatial approach followed the concept of video-CVEs, the other one was based on video streams in a physically fixed arrangement around an interactive table.

The study considered several subjective responses to assess and compare the user experience between conditions. First, it applied the well-proven social presence measure to assess the "salience of the other in the interaction". Second, it measured a feeling of spatial copresence. Third, it applied some specific questions that were informed by experiences gathered during previous experiments. These questions targeted aspects of interpersonal awareness and the ease of use of the systems. Participants were also asked to give a subjective preference ranking of the four conditions.

In addition, aspects of participants' communication patterns were compared between conditions to find out which interaction patterns of mediated collaboration best matched face-to-face collaboration. This was done using video-analysis.

6.1 *Experiment design*

The experiment was conducted by the author at the University of Canterbury in February and March 2006.

6.1.1 Experimental task

To obtain meaningful results on collaborative behaviour, the design of an appropriate task is crucial. To provoke a rich communication between participants to reveal the limits of different video-conferencing systems, a judgemental task was designed with highly ambiguous content. This follows from Media Richness Theory (see Section 2.2.1), in which more communication cues are required to resolve tasks with a high level of uncertainty.

In this case the task required participants to work together matching photographs of dogs to photographs of their owners. Participants were told during the introduction that a incidental result of this experiment should reveal if a study conducted by Roy and Christenfeld (2004), which showed that dogs and their owners resemble each other, could be replicated successfully for local dogs and owners.

In each of four rounds, a set of four photos of owners and four photos of their dogs were presented in random arrangements (see Figure 6.1 for an example[1]). The participants' task was to find the correct matches by discussing which dog might resemble which owner the most. Each team was allowed to take as much time as they needed to come up with an answer that both team members agreed upon, but they were also encouraged to take as little time as possible.

The photographs were taken especially for this experiment by the author, with consent of all dog owners. They were assured that their portraits would not appear in any publication that might result from this study. Therefore, the black bars are used in Figure 6.1 so as not to disclose their faces. All pictures of the owners showed the face of the person, the pictures of the dog showed either a portrait or a full body perspective of the dog, depending on its size. Out of a total of 30 pairs, five sets of four dog and owner pairs each

[1] Solution: A-III, B-IV, C-II, D-I.

Figure 6.1: An example set of four owners and their dogs that needed to be matched.

were formed with an equal balance of female and male owners, as well as a mixture of different dog breeds.

6.1.2 Experiment conditions

The experiment used a one-factor, repeated measures design, comparing different variables of the communication and collaboration across four conditions. The order of conditions was controlled in each experiment following a Latin square scheme.

The following four collaborative interfaces were implemented. All were suitable for a "photoware" task, where participants have to talk about, point at, move, and rotate digital or real pictures on a virtual or real table:

	FtF	**vTAB**	**sVC**	**vCVE**
Gaze supported	Yes	Yes	No	Yes
Table Interaction	Yes	Yes	No	Yes
Input	Gesture	Gesture	Mouse	Mouse
User View	Individual	Individual	Shared	Individual

Table 6.1: Main differences of the conditions in the experiment "Dogs & Owners".

(a) Face-to-Face (FtF)

(b) Video-table (vTAB)

(c) Video-conferencing (sVC)

(d) Video-CVE (vCVE)

Figure 6.2: Conditions in the experiment "Dogs & Owners".

1. Condition "Face-to-Face" (FtF): Unmediated face-to-face collaboration around printed photographs placed on a real table (see Figure 6.2(a)).

2. Condition "Video-Table" (vTAB): Mediated remote collaboration around a shared interactive table (see Figure 6.2(b)). Spatiality aspects were supported within a local, real-world reference frame. The digital photos were displayed and pre-arranged on a touch sensitive table surface that allowed for interaction with the pictures.

3. Condition "Standard video-conferencing" (sVC): Mediated collaboration through a standard WYSIWIS video-conferencing interface (see Figure 6.2(c)). No aspects of a shared three-dimensional reference

frame were given. This setup used a state-of-the-art video-conferencing system involving video streams of both participants displayed on the screen as well as a shared application window.

4. Condition "video-CVE"(vCVE): Mediated collaboration around a virtual table in a video-CVE (see Figure 6.2(d)). Spatiality aspects were supported within the remote, virtual space. While the interaction with digital photos was done with a standard computer mouse, the representations of the table and of the participants' video streams are shown in the simulated three-dimensional space. A special head-tracking device was used to allow for consistent virtual head-movement within the virtual environment.

Table 6.1 outlines the main differences between the conditions. These include whether it was possible for the users to have their individual spatial perspective onto the pictures, the spatial reference frame provided, whether digital or printed media were used, and what form of interaction was applied.

6.1.3 Apparatus

FtF: In this condition, both participants collaboratively examined a set of paper photographs in the same room while sitting on two opposite sides of a table (Figure 6.2(a)). The photos were of a standard format (13 cm×18 cm, resolution 1024×1280 pixels).

vTAB: In this setup, each participant was seated in front of a horizontally aligned, touch sensitive panel[2] which in turn was placed in front of a LCD monitor (Figure 6.2(b)). A projector under the table projected the photo application onto the touch panel. With a single finger, photos could be moved across the panel or rotated by dragging the rotation handles of a selected photo. The LCD monitor behind the touch panel showed the live video of the remote person. That person was seated in front of an identical setup, but with an upside-down version of the photo application running on the touch screen. Both participants had a clear idea of "their" side of

[2] http://www.nextwindow.com/, last accessed October 2007.

the panel and had their own individual view of the table. Half the photos were initially placed facing participant 1, and the other half facing towards participant 2 (upside-down for participant 1).

sVC: In this condition, the video-conferencing system Conference XP[3] version 4.0 was used. Two video windows were placed at the top segment of the LCD screen, one showing the participant's own video and one showing the other person's video. A shared photo application window was positioned underneath (see Figure 6.2(c)). Both participants could interact with the photos at the same time using a simple mouse click and drag interface. At all times, both users saw exactly the same content on the screen, as it is the case in most conventional video-conferencing tools. Photos that were uploaded at the beginning of the trial were all facing the same way (upright).

vCVE: In this condition, participants met in a virtual 3D room, represented as video-avatars around a virtual table, on top of which was running a shared photo-application (see Figure 6.2(d)). The interface was implemented using the cAR/PE! virtual tele-collaboration space (see Section 3.1.1). The head orientation of the participants was tracked with the commercially available 2DOF infrared tracker "TrackIR"[4] by Natural Point. Head tracking data was used to control the individual view into the virtual room. That way, person A could for example change his or her viewpoint between the table and the video avatar of person B by moving his or her head up and down. At the same time, the orientation of person A's video-avatar consistently followed the head movements, allowing person B in turn to infer what was in the view field, and thus the point of attention, of person A. The positions of the avatars were locked to the opposite sides of the table. Half the photos were flipped in the initial layout, so that half the photos could be seen in the correct orientation by each participant. To manipulate the photos, both participants used a standard mouse that controlled the shared mouse pointer displayed on the virtual table.

[3] http://research.microsoft.com/conferencexp/, last accessed October 2007.

[4] http://www.naturalpoint.com/trackir/, last accessed October 2007.

Audio and video recordings were made of the subjects using two DV-cameras with external microphones that were placed close to the participants. For all mediated conditions, two visually and acoustically separated rooms were prepared with identical standard desktop PCs (Pentium 4, 2.80 GHz), monitors (LCD, 17", 1280×1024 pixel resolution), headsets (stereo with mono microphone) and webcams (USB, CIF resolution). All computers involved in the setup were connected through a 100 Mb network switch.

The shared photo viewing application was based on the open source graphics editor Inkscape[5] version 0.43. Shared access to the application was implemented using the desktop sharing software UltraVNC[6] version 10.1.1.3. Both participants shared the same mouse pointer with equal manipulation privileges. The photo application as well as the UltraVNC Server and UltraVNC Client ran on two extra laptop computers which were also connected through the network switch. To capture the activity on the shared Inkscape window, one further PC was connected to the network switch which ran another UltraVNC client window that was captured in real time by the screen capturing software Camtasia[7].

6.1.4 Participants

Thirty volunteers (22 male and 8 female) participated in the experiment. In 15 sessions, teams of two took part in four trials for a total of 120 trials. The age of the participants ranged from 22 to 45 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. The participants were recruited from among post-graduate students and staff members from different departments at the local university. To exclude mixed gender effects and to make sure that all team members already knew each other before the experiment, every participant was asked to bring along a same-gender friend as his or her team partner.

[5] http://www.inkscape.org/, last accessed October 2007.

[6] http://www.uvnc.com/, last accessed October 2007.

[7] http://www.techsmith.com/camtasia.asp, last accessed in October 2007.

6.1.5 Procedure

For every one-hour session a group of two subjects was present. Upon arrival the participants were given the participant information sheet, which outlined (1) the goal of the experiment, (2) the general procedure, (3) the anonymity of the experiment, and (4) a participant consent form, which was to be signed by them. Additionally, the document contained a general demographics questionnaire.

A second sheet was handed out, describing the task according to Section 6.1.1. After questions regarding the task description were answered, each participant took part in four rounds, one round for each condition (FtF, vTAB, sVC, vCVE). The order of conditions was controlled beforehand following a Latin Square scheme. The task in each condition was the same. However, new sets of photos with different dogs and owners were used in each round.

In each video-conferencing condition, participants were given instructions on the use of the interface using the same special set of photos of dogs and owners that was shown on the photo application window during every warm-up phase. In the standard video-conferencing condition (sVC), participants were explicitly made aware that the other person sees exactly the same view as them at all times. In contrast, in the two spatial video-conferencing conditions (vTAB, vCVE), the individual view aspect of the interface was emphasised and the ability to infer the other person's gaze direction was pointed out. No instructions on the general strategy of how to find matching pairs were given.

In all three mediated conditions, the subjects wore audio head-sets which were explained and adjusted for best comfort. The head tracking in the vCVE condition was adjusted individually for every participant, so that all parts of the virtual table and the other participant's video-avatar could be viewed within a comfortable range of head positions.

Once both participants signalled that they had understood the interface and how to use it, a set of the actual experiment photos was loaded onto the shared photo application, indicating the official start of that round. It was now up to the participants to discuss and manipulate all the pictures that were on display and to identify the possible pairs. Suggested pairs could be indicated simply by moving a photo of a dog close to the photo of an owner. Once the team found four pairs that both team members explicitly expressed they agreed with, the round was finished.

Subjects were then brought back to the same room and were asked to fill out the experiment questionnaire. After that, the number of correct dog-owner pairs found in the last round was given to the team. After the fourth and final round was over, and the fourth questionnaire was filled out by the participants, they were briefly interviewed about how they liked the task and were asked to give their personal preference ranking of all four conditions they had just collaborated with. At the end of the experiment, the participants were thanked, and chocolate was given to them as a reward.

6.2 Measures

The measures used are explained in the following sections.

6.2.1 Social presence

Social presence was measured with a semantic differential scale as suggested in Short et al. (1976). In total, eight bipolar pairs were included in the experiment questionnaire (see Appendix A.3, Items Q19–Q26). Participants were asked to rate the communication media on a seven point scale according to pairs such as cold – warm, insensitive – sensitive, impersonal – personal, or unsociable – sociable.

6.2.2 Copresence

Four items in the experiment questionnaire addressed a perceived sense of spatial copresence (see Appendix A.3, Items Q2, Q10, Q11, Q16). The construct "Copresence" originated from the factor "Realism" which was extracted in the experiment "Desert Survival Game" (see Section 4.3.3).

6.2.3 Usability parameters

A further eleven items addressed different aspects of the usability of the system (see Appendix A.3, items Q1–Q15, excluding items Q2, Q10, and Q11).

6.2.4 Video recordings

Two DV cameras captured the audio and video of the outside view at each of the two workspaces. Additionally, the shared photo application window was recorded as a video by screen-capturing software. All three videos were synchronised and were rendered into a single video, which was subjected to the video analysis. The audio streams of the two participants were assigned to the left and right audio channel in the combined video. This was done with the video editing package Adobe Premiere Professional 1.5.

6.3 Data analysis

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

6.3.1 Social presence

Average factor scores for social presence were calculated and tested for an effect across the four conditions in an analysis of variance with “Medium” as a within-subject factor. Conditions were also compared pair-wise using the Bonferroni adjustment for multiple comparisons in case a significant main effect was found.

6.3.2 Copresence

In the same manner as social presence, copresence factor averages were tested for an effect across the four conditions in an analysis of variance with “Medium” as a within-subject factor. Conditions were also compared pair-wise using the Bonferroni adjustment for multiple comparisons in case a significant main effect was found.

6.3.3 Usability parameters

Since the usability items were not expected to correlate, an analysis of variance with "Medium" as within-subjects factor was conducted for every item individually. Furthermore, post-hoc pairwise comparisons were calculated in case a main effect was found before.

6.3.4 Video analysis

The combined video files were analysed with respect to four communication quality parameters: (1) task completion time, (2) turns per minute, (3) technology and process versus task related turns, and (4) deictic versus descriptive references.

6.3.5 Preference

Participants were asked to rank the conditions from 1 (most preferred condition) to 4 (least preferred condition) according to their personal experience during the study. Significant differences between the average rankings were assessed with a non-parametric Friedman test.

6.3.6 Scale validity

The constructs "social presence" and "copresence" were subjected to a confirmatory principle component factor analysis (for more details refer to Section 4.2) to test for their uni-dimensionality. They were also tested for internal consistency by the determination of Cronbach's alpha.

6.3.7 Gender effects

Average factor scores of social presence and copresence were also subjected to a 4×2 mixed-factor ANOVA with the within-subjects factor "Medium" and the between-subjects factor "Gender".

This test was purely exploratory, mainly to investigate whether there was an interaction between the gender of the teams and the medium used. In this experiment, the gender of the teams was controlled for the first time, so that all the teams were either male-male or female-female.

6.4 Results

A total of 15 sessions were run with 2 participants in each session, where the first two sessions were initial pilot trials whose results were not considered in this statistical analysis. Therefore, 13 sessions form the basis the results reported below. All questionnaires of the 26 subjects were valid. No values were missing.

All participants (except one self-described "cat person") liked the task and quickly became engaged in finding the matching pairs. The most common judgement criteria were whether a dog would belong to a woman or a man, if a dog would match the more active or passive lifestyle inferred from the photos of the owners, and matching hair colour and facial features between owners and dogs. For the total of 16 dogs presented in each experiment, on average 5.2 correct owners were found. This is slightly more than would be expected in a total random scenario and could indicate that the participants had some slight ability to match dogs with their owners[8].

The teams' strategy of handling the photograph orientation was consistent over all three conditions that involved individual viewpoints. Two main strategies were used to either rotate all pictures to be correctly oriented for person A first, and then rotate them all back so person B could have a look; or, to place the photos in the middle of the table and rotate them about 90 degrees into a more neutral sideways position where both could examine them at the same time. Occasionally, in the condition vTAB, the participants had difficulty rotating photos using the touch-sensitive table due to problems acquiring the rotation handle.

6.4.1 Social presence

The average factor scores and standard errors for social presence in the different conditions are shown in Figure 6.3. There was a significant main effect between the four different media, $F(3,75)=21$, $p<0.001$. Post-hoc comparisons showed that social presence was significantly higher in FtF ($M=6.4$, $SD=0.6$) than vTAB ($M=4.8$, $SD=0.9$, $p<0.001$), sVC ($M=4.5$, $SD=1.3$,

[8] Mentioned here for completeness only. This result was not investigated further as it was outside the focus of this study.

p<0.001), and vCVE (M=4.9, SD=1.2, p<0.001). However, none of the mediated conditions showed significant differences in pair-wise comparisons. Although not significant, the average of social presence was rated higher in both spatial conditions than in the sVC condition.

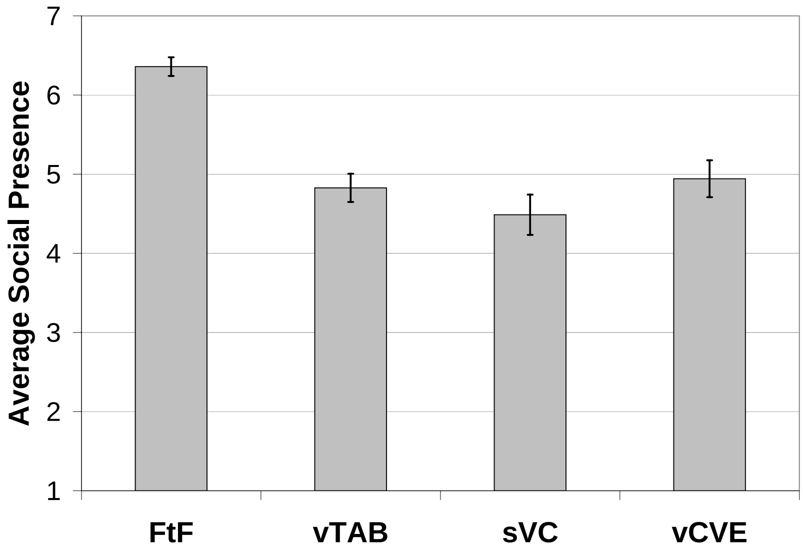


Figure 6.3: Social presence averages and standard errors.

6.4.2 Copresence

The average factor scores and standard errors for copresence in the different conditions are shown in Figure 6.4. There was a significant main effect between the four different media, F(3,75)=64, p<0.001. Post-hoc comparisons showed that participants felt a higher sense of spatial copresence in FtF (M=6.7, SD=0.5) than vTAB (M=4.3, SD=1.1, p<0.001), sVC (M=3.5, SD=1.1, p<0.001), and vCVE (4.2, SD=1.1, p<0.001). Furthermore, the perceived copresence in vTAB was significantly higher than in sVC (p=0.043).

6.4.3 Usability parameters

Eleven items addressed different aspects of the usability of the system. As these items were not expected to measure a single construct, the results were

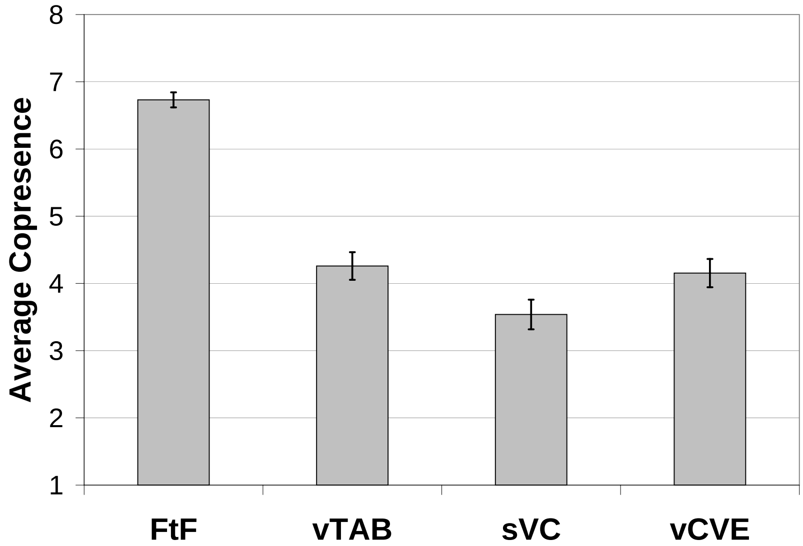


Figure 6.4: Copresence average ratings and standard errors.

calculated for every item individually. The questions and their scores are shown in Table 6.2.

Except for questions 5, 7, and 8, all results showed a significant main effect. Post-hoc comparisons revealed that many of these effects reside in the large difference of the scores between the face-to-face and the mediated conditions. However, two significant differences between the spatial and the 2D videoconferencing interface could be found. The score for question 4, "I could easily tell where my partner was looking" was significantly higher in condition vTAB than in the condition sVC (p=0.02), and also significantly higher in the condition vCVE than in the condition sVC (p=0.03). Furthermore, the results of question 6, "I was often confused", uncovered that participants felt confused more often in the condition vTAB than in the condition sVC (p=0.05). They also felt more often confused in the condition vCVE than in the condition sVC (p=0.05). The results in all other usability and communication related items show condition sVC to be closer to FtF than both spatial video-conferencing conditions.

Question	FtF	vTAB	sVC	vCVE	Post-hoc comparisons
1) It was very easy to make myself understood.	6.4 (1.2)	5.4 (1.1)	5.9 (1.2)	4.9 (1.6)	FtF>vTAB; FtF> vCVE
2) I could easily tell where my partner was pointing at.	6.7 (0.7)	4.9 (2)	4.3 (2.2)	4.7 (1.8)	FtF>vTAB; FtF>sVC; FtF> vCVE
3) I could not contribute anything to the solution we came up with.	1.6 (0.7)	2.1 (1.2)	1.8 (0.8)	2.2 (1.1)	FtF<vCVE
4) I could easily tell where my partner was looking.	5.8 (1.5)	4.6 (1.7)	2.9 (1.7)	4.2 (2.0)	FtF>vCVE; FtF>sVC; vTAB>sVC; vCVE>sVC
5) There was a lot of time when no-one spoke at all.	2.4 (1.6)	3.3 (1.7)	2.6 (1.5)	3 (1.7)	*
6) I was often confused.	1.7 (1.0)	3.1 (1.8)	2.1 (1.2)	3.5 (1.9)	FtF<vTAB; FtF<vCVE; sVC<vTAB; sVC<vCVE
7) We were never talking over one another.	5 (2.0)	4.4 (1.6)	4.2 (1.6)	4.4 (1.6)	*
8) I hardly looked at my partners face.	4 (2.2)	3.3 (1.7)	4.5 (2.1)	3.9 (1.9)	*
9) I knew exactly when it was my turn to speak.	5.8 (1.0)	4.5 (1.5)	4.9 (1.5)	4.7 (1.4)	FtF>vTAB; FtF>sVC; FtF>vCVE
10) I could always clearly hear my partners voice.	6.7 (1.0)	5.2 (1.6)	5.9 (1.1)	5.6 (1.5)	FtF>vTAB; FtF>sVC; FtF>vCVE
11) When I looked at my partner, I could always clearly see his or her face.	6.8 (0.4)	5.5 (1.4)	5.3 (1.9)	5.9 (1.0)	FtF>vTAB; FtF>sVC; FtF>vCVE

*Note: Standard deviation in parentheses, * = no significant differences*

Table 6.2: Average scores and standard deviations for the eleven usability questions in the questionnaires on a 7-point, Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree).

6.4.4 Preference

Every participant ranked FtF first. To find out which of the three mediated conditions was the second most preferred, the rankings of the other three conditions were subjected to a Friedman test. The average raking of sVC (M=2.6) was lower than both vTAB (M=3.2) and vCVE (M=3.2), indicating that the standard video-conferencing interface was marginally preferred over the spatial interfaces. This difference, however, did not reach significance (Friedman test, χ_r^2= 5.8, df=2, N=26, p=0.056).

Variable	FtF	vTAB	sVC	vCVE	Post-hoc comparisons
1) Task completion time (sec)	192 (131)	306 (165)	163 (59)	414 (206)	vCVE>FtF; vCVE>sVC
2) Total turns per minute	5.4 (2.3)	4.2 (0.5)	5.0 (1.7)	4.1 (1.7)	*
3) Technology and prosess related turns out of total turns	0.12 (0.10)	0.26 (0.08)	0.12 (0.08)	0.4 (0.18)	vCVE>sVC; vCVE>FtF; vTAB>sVC
4) Ratio deictic references to total references.	0.98 (0.04)	0.78 (0.17)	0.7 (0.25)	0.65 (0.20)	FtF>vTAB; FtF>sVC FtF>vCVE

*Note: Standard deviation in parentheses, * = no significant differences*

Table 6.3: Averages and standard deviation of video analysis parameters.

6.4.5 Video analysis

The video observation analysis was done by the author. Due to technical difficulties only 12 out of 13 videos were completely captured and available for analysis. Table 6.3 summarises the results of the video analysis.

Task completion time

The task completion time was defined from the moment when the participants first saw the photos of the dogs and owners until the moment when they explicitly signalled that they found a solution both agreed with.

Results varied significantly across the four conditions, $F(3,33)=9.1$, $p<0.01$, where condition sVC was the fastest, followed by condition FtF, condition vTAB and at the end, taking on average more than twice as long as condition sVC, condition vCVE. Post-hoc analysis found significant differences between conditions vCVE and FtF, and between conditions vCVE and sVC.

Turns per minute

The spoken turns of both participants were counted during the video analysis. The same definition of a turn as in Sellen (1995) was used according to which "a turn consists of a sequence of talk spurts and pauses by a speaker that holds the floor." During the video analysis, turns were counted for one person at a time and the number of turns of both participants was then summed to determine the total turns. As the absolute number of turns would not be comparable to other conditions due to the different durations of the rounds, the number of total turns was divided by the task completion time. The so gained value of total turns per minute can be considered as a variable that indicates the dynamics of the communication flow.

FtF and sVC had slightly more turns per minute on average, suggesting a more vivid communication flow. However, these differences did not reach significance in the test for the main effect.

Turn content

The content of each turn was examined to see if it related either to the collaborative task, or if it related to the use of the technology involved or the collaborative process. For example the content of the statement: "I think this dog doesn't look at all like this guy" is task related, whereas statements like "Did you just move your mouse" or, "I think you should first rotate the dogs so you can see them, and then I will do the same afterwards" fit more in the technology or process related category.

By constructing the ratio of all the non-task related turns by the total number of turns, an indicator of the extent that the technology "got in the way" during the collaboration was obtained. The result showed a significant main effect across the four conditions, $F(3,33)=18$, $p<0.01$. Post-hoc comparisons revealed that the occurrence of non-task related turns was significantly higher in the condition vTAB than in the conditions FtF ($p=0.01$) and sVC ($p=0.03$). The occurrence of non-task-related turns was also found to be higher in the condition vCVE than in conditions FtF ($p<0.01$) and sVC ($p=0.03$).

Deictic references versus descriptive references

As explained in Section 2.1.3, deictic references are the easiest, fastest, and most concise way of referencing. However, they are less frequently used in mediated communication, since it is harder to maintain a shared referencing context with the absence of certain visual communication cues. By forming the ratio of deictic to descriptive references an indicator for communication efficiency can be obtained and compared across the four media conditions.

In all 12 videos, all references to either dogs or owners were registered during the video analysis and were counted either as deictic ("that dog", "him","her", "that guy") or as descriptive ("the girl with the glasses", "the Labrador", "the third dog from the left"). Out of the total number of references, the ratio of deictic references was calculated and compared between all conditions. A significant main effect was found, $F(3,33)=18$, $p<0.01$. Further post-hoc analysis showed that the relative occurrence of deictic references out of all registered references was significantly higher in FtF than in conditions vTAB ($p<0.01$), sVC ($p=0.01$), and vCVE ($p<0.01$).

6.4.6 Scale validity

The principle components factor analyses extracted one factor for social presence (explaining 58% of the variance), and one factor for copresence (explaining 69% of the variance). The uni-dimensionality of the two constructs was therefore confirmed.

Two items of the social presence measure were excluded from analysis because of their lowest factor loadings. These items were "small–large" (factor loading 0.56) and "spontaneous–formal" (factor loading 0.64). Internal consistency for the remaining six bipolar pairs was high (α=0.90).

All four copresence items had factor loadings higher than 0.7 and were retained. Internal consistency for those four items was sufficient (α=0.84).

6.4.7 Gender effects

There was a significant interaction Medium×Gender, $F(3,72)=4.1$, $p=0.009$, for social presence. As can be seen in Figure 6.5, female participants rated social presence of the standard video-conferencing condition higher than the male participants, while the three remaining conditions did not show any salient differences.

No significant interaction Medium×Gender was observed for copresence, $F(3,72)=2.4$, $p=0.08$.

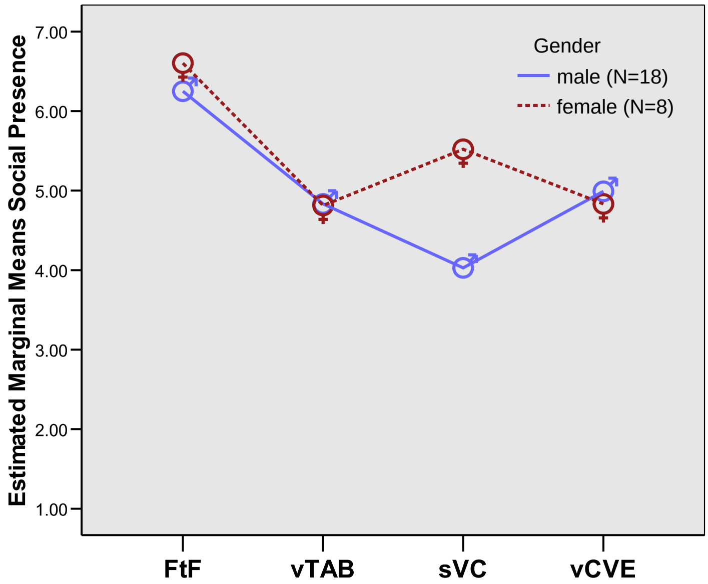


Figure 6.5: Interaction between conditions and gender for social presence.

6.5 Discussion

The hypothesis that adding spatiality fosters a feeling of being face-to-face was supported by higher social presence, copresence, and gaze awareness scores for the spatial interfaces.

However, the hypothesis that adding spatiality supports face-to-face-like behaviour was rejected. As it turned out, the spatial context introduced additional mental effort caused by the more complex interaction and usability problems, which lead to confusion and distraction from the task. The collaborative behaviour and communication was therefore subject to interference caused by this cognitive overhead, and thus turned out to be less efficient and less task-focused, and thus less face-to-face-like, than the collaborative behaviour afforded by the simple standard video-conferencing interface.

Adding spatiality therefore involves a tradeoff: spatial VMC systems are capable of creating a collaborative context that is closer to face-to-face, but at the same time lose the efficiency of a task focused two-dimensional interface. In this experiment, that trade did not pay off as indicated by the low preference scores of the spatial interfaces.

The gender difference in social presence for condition sVC was unexpected and is hard to interpret ex post because the study did initially not focus on gender effects, and thus important hints into what caused the differences may have slipped through the net. However, in an attempt to provide an interpretation based on the data that was gathered, all results for condition sVC were reviewed for salient differences between male and female participants which could help understand what may have caused the differences in social presence in that condition. One difference was found in condition sVC for the questionnaire item Q15: "I could easily tell where my partner was looking". Female participants (M=4, SD=1.8) agreed to this statement significantly more than male participants (M=2.4, SD=1.5), as tested in an independent samples t-test, $t(24)= -2.3$, p=0.028. Since the videos did not provide visual information about gaze direction, one explanation for this difference could be that female teams better informed each other *verbally* about where they were looking during the collaboration. Consequently, easier-to-use communication systems that demand more verbal communication may convey more social presence for female participants, while more complex systems that afford spatial cues such as gaze awareness may convey more social presence for male participants.

Despite the fact that this reasoning remains speculative at this point, the observed gender effects observed in this experiment brought up the question if sex-differences communication styles may influence the subjective perception of a communications medium, which suggests that gender should be considered a human factor in experiments comparing VMC systems.

6.6 Experiment summary

This chapter presented the design and findings of a study comparing two video-conferencing interfaces that support spatial cues with a standard video-conferencing system, as well as with a same room face-to-face condition. Various differences between the conditions were found which suggest that the spatial character of an interface can support a feeling of being face-to-face, manifested in higher degree of gaze awareness, social presence, and copresence, while at the same time compromising the task focus and efficiency of a two-dimensional interface which approximates the simplicity of face-to-face interaction. Participants in the experiment preferred the more efficient interface over the two spatial interfaces that supported a more face-to-face-like context. Differences in responses were found for male and female teams, which suggests that gender should be considered a factor when studying collaboration with different VMC interfaces.

The next chapter investigates the impact of the level of transportation of a video-CVE on user experience and collaboration, and further explores occurrences and causes of gender effects.

Chapter 7

Experiment: "Celebrity Quotes"

This chapter presents the "Celebrity Quotes" experiment, the fourth and final study. This experiment focused on the impact of the level of transportation of a video-CVE on the user experience and collaborative behaviour. The transportation property of a communication medium is determined by the extent to which its interface is designed to "mentally transport" the participants from their physical environment to a remote mediated context (see Section 3.2.2). The level of transportation of a communication medium is therefore directly linked to the experience of presence at a remote space.

The hypothesis this experiment investigated was that more "transporting" video-CVEs induce a stronger sense of physical presence and afford a user experience and collaborative behaviour that is closer to being face-to-face. The study "Celebrity Quotes" thus extended the findings of the experiment "Dream House" (Chapter 5) which explored the correlation between a sense of physical presence and social presence of standard video-conferencing systems and video-CVEs, and complemented the experiment "Dogs & Owners" (Chapter 6) which studied the impact of spatiality of a VMC interface.

The introduction of the experiment "Dogs & Owners" mentioned the problem of separating the impact of spatiality from the impact of transportation of a video-CVE. This experiment sought to isolate the impact of transportation of a video-CVE on user experience and user behaviour by the selection of a suitable scenario, and the inclusion of two video-CVE conditions with different levels of transportation, but same levels of spatiality. The scenario chosen involved a task and a virtual room, which was designed to foster a sense of "being there"(see Section 7.1.2). The conditions included two video-CVEs, a standard video-conferencing system, and being face-to-face. The level of transportation of the two video-CVEs was controlled by the deployment of more or less immersive displays: the less "remote" video-CVE was displayed on a regular-sized desktop computer screen; the more "remote" video-CVE was displayed as a large stereoscopic projection. Based on Nash et al. (2000, page 32) and Sadowski and Stannay (2002, page 796) the larger field of view and stereopsis of the more "remote" video-CVE was expected to induce a stronger sense of physical presence.

The experiment assessed physical presence, social presence, copresence, awareness, and preference. It also included a series of questions assessing the ease of use of the system which were informed by the usability issues that surfaced in the experiment "Dogs & Owners".

Furthermore, extracted conversation transcripts of participants' dialogues were analysed and compared to find out which communication patterns in mediated conditions matched the most with patterns obtained in the face-to-face condition. The type of content analysis was motivated by Kramer et al. (2006), who reported an effect of the level of perceived physical presence on the occurrence of utterances of certain linguistic features. Kramer et al.'s (2006) basic premise was that a sense of "being there" in a remote space correlates with the extent to which people talk about a remote space in the same way they talk about local space. Building on this premise, differences in linguistic patterns were therefore also expected to emerge between the two video-CVE conditions which induced different levels of physical presence. Complementarity to the transcript analysis, session videos were also analysed with regards to the way participants coordinated their views of the shared workspace.

To further explore the persistence and causes of gender effects like the one that surfaced in experiment "Dogs & Owners", more attention to differences that occurred between male and female teams was given a priori, leading to a 4×2 (Medium × Gender) multi-factorial data analysis.

7.1 Experiment design

This experiment was conducted by the author at the University of Canterbury in March and April 2007.

7.1.1 Experiment task

A collaborative pair-matching task was designed for this experiment. Ten photos of well-known personalities as well as ten significant quotes were presented to each team in each of the four rounds (see Figure 7.1 for an example[1]). The task for the team in each round was to find as many correct celebrity–quote pairs as possible.

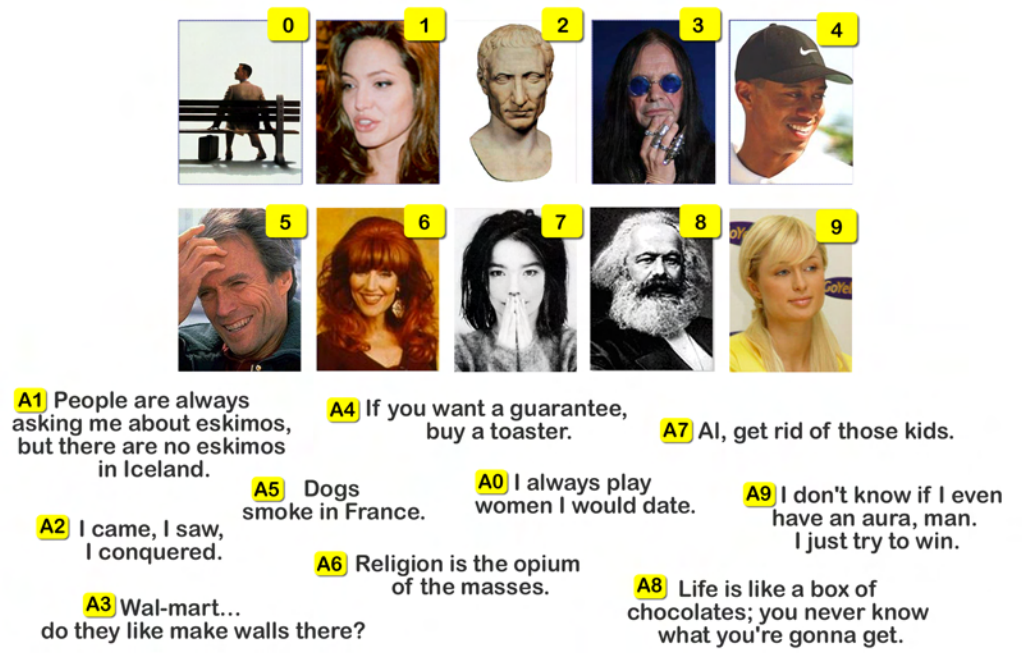


Figure 7.1: An example set of quotes and photos. Participants had to collaborate to find matching pairs.

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, actors, athletes, and fictional characters. That way it was certain that every participant would know at least some of them, irrespective of their background. Based on findings of a pilot study by the author involving 60 participants, celebrities for each set were also combined so as to equalise the challenge among the sets as much as possible, that is,

[1] Solution: 0–A8, 1–A0, 2–A2, 3–A5, 4–A9, 5–A4, 6–A7, 7–A1, 8–A6, 9–A3.

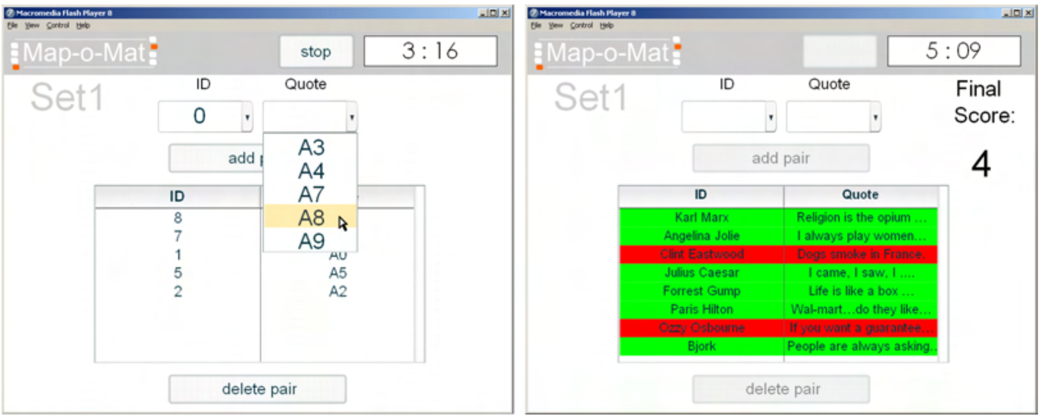


Figure 7.2: The "Map-o-Mat" mapping application. It allowed participants to enter celebrity–quote pairs.

every set contained the same number of quote–celebrity pairs that were found to be "hard", "medium", or "easy".

The photos portrayed the celebrities in a "typical" way, so that participants would be able to recognise them. The photos were all collected from various publicly accessible Internet sites. The quotes were all retrieved from an Internet database for famous quotes[2]. Every photo had its own ID, every quote had its own quote-tag, as displayed in the little yellow boxes (see Figure 7.1).

Participants could enter every celebrity–quote pair they found into a special application, the "Map-o-Mat" (see Figure 7.2), which was developed by the author for this study. Entering pairs was done by selecting the ID of a celebrity as well as the corresponding quote-tag from a pull-down menus associated with every quote and click on a "add pair" button (see Figure 7.2, left). Once the members of a team decided to stop putting in more pairs, they could press a stop button which immediately processed the entered items and displayed the results of that round (see Figure 7.2, right). The final score was calculated as the number of correct pairs minus the number of incorrect ones. Participants therefore also had to ponder how much risk they wanted to take when entering pairs they were not sure about.

[2] http://www.brainyquote.com/, last accessed in October 2007.

As in the other experiments, the main focus of this experiment was not the task outcome, but the collaborative process. Any forms of learning effect that could have influenced the task score were therefore of minor concern.

7.1.2 Experiment scenario

The chosen collaborative scenario was to present the three artefacts ("photos", "quotes", and "Map-o-Mat" application) separately, so that they could not be looked at the same time, and would involve some form of effort to switch from one to the other. The design of all conditions was based on this scenario.

The rationale for this was that firstly a spatial separation of the artefacts created the need for navigation in the remote video-CVE conditions. The experience of "self-motion" in a remote, virtual environment is a pivotal contributing factor for a user's sense of actually "being there" and thus underlines the transportation aspect of these interfaces.

Secondly, a separation of the artefacts allowed exploration of the different ways that participants coordinated their collaboration in the absence and presence of different visual awareness cues. The effort to switch from one artefact to the other was expected to foster the emergence of clearer collaborative strategies following the least effort principle.

Figure 7.3 illustrates the scenario in the example of the virtual room used in the video-CVE conditions. The photos and the quotes are presented on two spatially separated billboards. The Map-o-Mat terminal was located in between the billboards. The size of each artefact was chosen to be small, so that participants had to navigate their avatar close to see them clearly.

The proximity of an avatar was, in turn, a visual awareness mechanism for the other participant, who could automatically infer what the other person was currently looking at based on the position and orientation of his or her avatar. In the example given in Figure 7.3, one of the participants is operating the Map-o-Mat application, while the second participant is reading the quotes on the billboard on the right. The avatars appear transparent in this illustration, since they are facing away from the viewer's perspective.

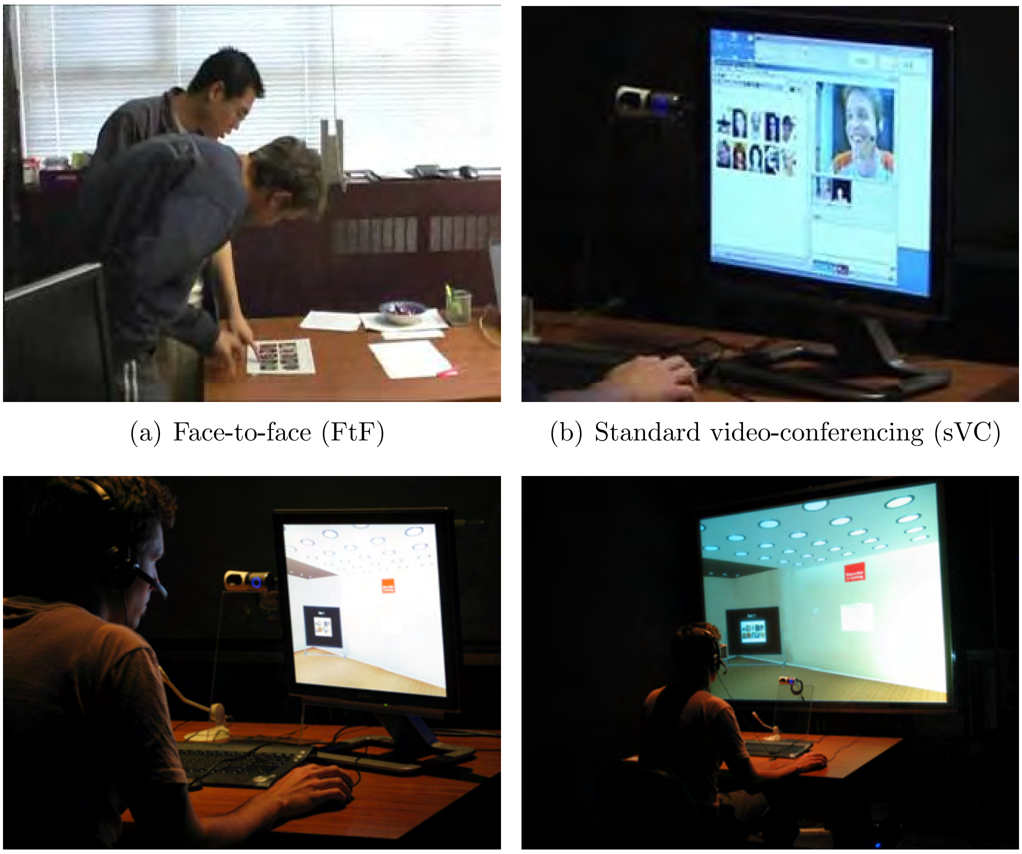
Figure 7.3: The virtual room used in the experiment. Two billboards show the quotes and photos; an integrated terminal shows the Map-o-Mat application. As can be seen by the position of the avatars in the figure, one participant is entering some data into the Map-o-Mat (centre), while the other participant is reading some quotes on the billboard (right).

7.1.3 Experiment conditions

The previously explained scenario was implemented in each of the following four conditions.

1. Condition “face-to-face” (FtF): Participants collaborated in the same room (see Figure 7.4(a)). Two sheets of paper showing photos and quotes were attached on the opposite ends of a long table 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 the table (not visible in the photo).

2. Condition “standard video-conferencing” (sVC): Participants were located in separate rooms and were connected via a commercially available VC-tool (Marratech) which showed the video streams of both participants and a shared presentation area where celebrities and quotes



(a) Face-to-face (FtF)

(b) Standard video-conferencing (sVC)

(c) Video-CVE desktop (vCVE_desk)

(d) Video-CVE immersive (vCVE_im)

Figure 7.4: Conditions in the experiment "Celebrity Quotes".

were displayed on different slides (see Figure 7.4(b)). Every 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 he or she was looking at the same slide in that moment.

To introduce the same effort to switch between the view showing quotes and photos as was given in the other conditions, twelve additional "buffer slides" were added between the photo-slide and the quotes-slide. Participants could flick through the slides using the *PageUp* and *PageDown* button. To change from the photo-slide to the quotes-slide therefore involved twelve button presses, which lasted approximately as long as it took people to navigate between the artefacts in the other conditions (three seconds).

The Map-o-Mat ran in a separate application window (in the background in Figure 7.4(b)). The application was shared, allowing both participants to control the application and enter pairs.

3. Condition "Video-CVE desktop"(vCVE_desk): Participants were located in separate rooms and met virtually in the previously illustrated shared virtual room. In this condition, the virtual room was presented to each of the participants on a standard 19" flat screen monitor placed on the desktop in front of them (see Figure 7.4(c)). Participants could control both the Map-o-Mat application and point at features on the billboards using a standard computer mouse. In addition, a commercially available space mouse was provided as an easy, state-of-the-art means for navigation within a virtual space. Participants typically concurrently operated the space mouse with their left hand and the normal computer mouse in the right hand. The time it took to navigate from the billboard showing the photos to the billboard showing the quotes was approximately three seconds.

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 a 80" stereo-projection (see Figure 7.4(d)). To create the stereoscopic effect, participants wore special shutter glasses.

7.1.4 Apparatus

FtF: The photos and quotes were colour-printed on A4 sheets of paper which were then laminated and attached to opposite ends of a 2.4 meters long table using Velcro strips. They were oriented sideways, so that participants needed to walk around the ends of the table in order to see them properly. A regular laptop computer (P4, 1.6 GHz) was used to run the Map-o-Mat application.

sVC: This condition was based on a full version of the video-conferencing software "Marratech 6.1", which was provided by the company Marratech[3] for this experiment. Marratech 6.1 was considered a practical and suitable state-of-the-art video-conferencing tool for the scope of this experiment, because it allowed the upload and the display of the prepared slides into a integrated whiteboard area, which allowed users to browse through these slides individually, following a relaxed WYSIWIS paradigm.

The Map-o-Mat application ran on a third computer which shared the application with the two participants' computers using the desktop-sharing software UltraVNC[4] version 10.1.1.3.

vCVE_desk: This condition was based on the video-CVE cAR/PE! (see Section 3.1.1). The virtual room used was customised for this experiment. Participants navigated their avatars using the space mouse "SpaceTraveler"[5]. The access to the integrated Map-o-Mat application (VNC client) was realised by using the open source software "Synergy"[6], which allowed one mouse pointer to be shared between the computer running the cAR/PE! application and computer running the Map-o-Mat application (VNC server). To jump between controlling the vCVE and the Map-o-Mat with their computer mouse, participants had to press two pre-configured extra buttons on the side of their mouse.

vCVE_im: The image of the virtual environment was projected with a DepthQ "InFocus" stereo projector[7], which, in combination with active LCD shutter glasses, is capable of supporting active stereo-vision. The shutter glasses used were the model "60GX" by NuVision[8]. The projector reached

[3] http://www.marratech.com/, last accessed October 2007.

[4] http://www.uvnc.com/, last accessed October 2007.

[5] http://www.3dconnexion.com/3dmouse/spacetraveler.php, last accessed November 2007.

[6] http://synergy2.sourceforge.net/, last accessed November 2007.

[7] http://www.depthq.com/projector.html, last accessed November 2007.

[8] http://www.vrs.com.au/stereoscopic/nuvision/60gx-shutter-glasses.html, last accessed November 2007.

an update rate of 120 Hz, which, after separating the images for left and right eye, resulted in an effective frame rate of 60 Hz. A stereo signal of the virtual scene was created by activating Open Inventor's built-in stereo capabilities. To prevent discomfort for inexperienced users, the camera offset (that is the virtual eye separation) in the stereo adjustment was set to a low 0.10.

The PCs used to run the video-conferencing software at both ends were two identical Shuttle PCs (AMD Athlon, 2.2 GHz, Dual Core), equipped with high end graphics cards (NVidia Quadro FX series). 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 (stereo audio, mono microphone). 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.

7.1.5 Participants

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. The participants were recruited among post-grad students from different departments at the University of Canterbury. To exclude mixed gender effects and to make sure that all team members already knew each other before the experiment, every participant was asked to bring along a same-gender friend as his or her team partner. Furthermore, participants were required to be fluent in English.

7.1.6 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.

The task was then explained to the participants using a printed set of photos and quotes which was not used in the experiment later on. After any questions were answered, the first of the 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. The task in each condition was the same. However, new sets of photos and quotes were used in each round.

In each video-conferencing condition, participants were given instructions on the use of the interface using the spare set of photos and quotes. In the first round that involved one of the two video-CVE conditions, the use of the SpaceTraveler was explained to the participants with a special navigation tutorial application provided by the manufacturer of the device.

In the standard video-conferencing condition (sVC), participants were explicitly made aware that the other person does not necessarily see the same slide in the Whiteboard.

In the two vCVE conditions, the individual-view aspect of the interface was emphasised and the ability to infer the other person's activities based on the location and orientation of his or her avatar was pointed out. In condition vCVE_im, the purpose of the shutter glasses along with the basic principle of stereo-vision was explained to the participants. After they put the glasses on, a quick informal stereo-vision test was conducted to confirm that the participant was capable of perceiving a sense of depth in the image and to make sure that it did not cause discomfort.

In all three mediated conditions, the subjects wore audio head-sets which were adjusted for best comfort. No advice on the general strategy how to find matching pairs was given.

The room with the actual sets of photos and quotes was loaded and the Map-o-Mat was started once both participants signalled that they understood the interface and felt confident using it. This officially started the round. It was now up to the participants to identify the celebrities on the photo board and discuss which of the quotes each of them might have said. Suggested pairs were entered 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.

Subjects were then brought back into the same room and were asked to fill out the experiment questionnaire. After the fourth and final round was over and the fourth questionnaire was filled out by the participants, they were briefly interviewed about their experience with the different interfaces and were asked to give their personal ranking of all four conditions they had just collaborated with. At the end of the experiment, the participants were thanked, and a coffee voucher was given to them as a reward.

7.2 Measures

The measures used are explained in the following sections.

7.2.1 Social presence

Social presence was measured using a semantic differential scale like that suggested in Short et al. (1976). In total eight bipolar pairs were included in the experiment questionnaire (see Appendix A.4, Items Q18–Q25). Participants were asked to rate the communication media on a seven point scale according to pairs such as cold – warm, insensitive – sensitive, impersonal – personal, or unsociable – sociable.

7.2.2 Physical presence

Six items in the experiment questionnaire addressed a perceived sense of physical presence (see Appendix A.4, Items Q12-Q19). The items were a sub-set of Schubert et al.'s (2001) Igroup Presence Questionnaire (IPQ).

7.2.3 Items on awareness, copresence, and ease of use

Eleven items were included in the experiment questionnaire which targeted spatial copresence (see Appendix A.4, Q3, Q5, Q6), Awareness (Items Q1, Q4, Q8, Q9), and Ease of Use (Items Q2, Q7, Q10, Q11).

7.2.4 Video recordings

Two DV cameras captured audio and video at each of the two workspaces. These videos were synchronised and were rendered into a single video which was analysed. The audio streams of the two participants were assigned to the left and right audio channel in the combined video. Video editing was done with Adobe Premiere Professional 1.5. From the synchronised videos, a stereo audio track (.wav format) was extracted with Apple Quicktime Pro 7.2 for later transcription.

7.2.5 Communication transcripts

Communication transcripts were manually created from the extracted audio files using the open source software "Transcriber"[9] (see Appendix C.1 for an excerpt of a transcribed conversation). The following rules were applied for the transcription process:

1. Complete quotes that were read aloud by any participant were not transcribed word-by-word to prevent the words in the quotes having an impact on the communication patterns that were meant to be compared across the conditions. Instead, they were marked as "QUOTE" in the transcript.

2. 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.

3. 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.

4. All occurrences of nonfluencies, like "Hm, hmm, uh, uhh, uhm, ahm, um, ..." were transcribed uniformly as "hmm"

[9] http://trans.sourceforge.net/en/presentation.php, last accessed November 2007.

5. Fillers, such as “you know”, “I mean”, and “I don't know” were transcribed as one word, “youknow”, “Imean”, and “Idontknow”

6. No comments of any type were included in writing by the transcriber, since they would affect later analysis of the transcripts.

Transcription files were exported as text files, from which the number of utterances and the overlapping utterances were determined first. Then a further two text files were extracted which included the transcript of each of the two participants separately.

7.3 Data analysis

The questionnaire data was analysed with SPSS version 14.0. The significance level was set to 0.05. Based on the indication of social presence being sensitive to gender effects (found in the experiment “Dogs & Owners”), all data in this experiment was analysed following the mixed factor design Medium(4)×Gender(2), with “Medium” as a within-subject factor, and “Gender” as a between-subjects factor. If a significant main effect was found between the four media conditions, pair-wise comparisons were performed using the Bonferroni adjustment for multiple comparisons.

7.3.1 Social presence

Average factor scores for social presence were calculated and tested for main effects and interactions across the four conditions in a mixed factor analysis of variance.

7.3.2 Physical presence

Average factor scores for physical presence were calculated for the two vCVE conditions and tested for main effects and interactions in a mixed factor analysis of variance.

7.3.3 *Copresence, awareness, and ease of use*

The eleven items targeting copresence, awareness, and ease of use were subjected to a principle component factor analysis to test the individual item's factor loadings and to confirm the dimensionality of the constructs.

Average factor scores were calculated for extracted factors and tested for main effects and interactions across the four conditions in a mixed factor analysis of variance.

7.3.4 *Preference*

Participants were asked to rank the conditions from 1 (most preferred condition) to 4 (least preferred condition) according to their personal experience during the study. Significant differences between the average rankings were assessed with a non-parametric Friedman test.

7.3.5 *Linguistic features analysis*

Extracted transcripts were analysed with the software "Linguistic Inquiry and Word Count (LIWC2001[10])" (Pennebaker et al. 2003). 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 C.1 in Appendix C.2 gives an overview of the test dimensions included in the dictionary. The result of a LIWC2001-analysis of a given text is an overview of standard linguistic dimensions such as the total number of words or the number of words per sentence, as well 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

[10] http://www.liwc.net/, last accessed November 2007.

across the four conditions in a mixed factor analysis of variance.

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.

7.3.6 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 only to the same subset of groups that were also 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" (Looser 2007) was used. This program allows the experimenter to open a video and press pre-defined buttons to record activity states of interest along a shared time line (see Figure 7.5).

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.

7.4 Results

7.4.1 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.

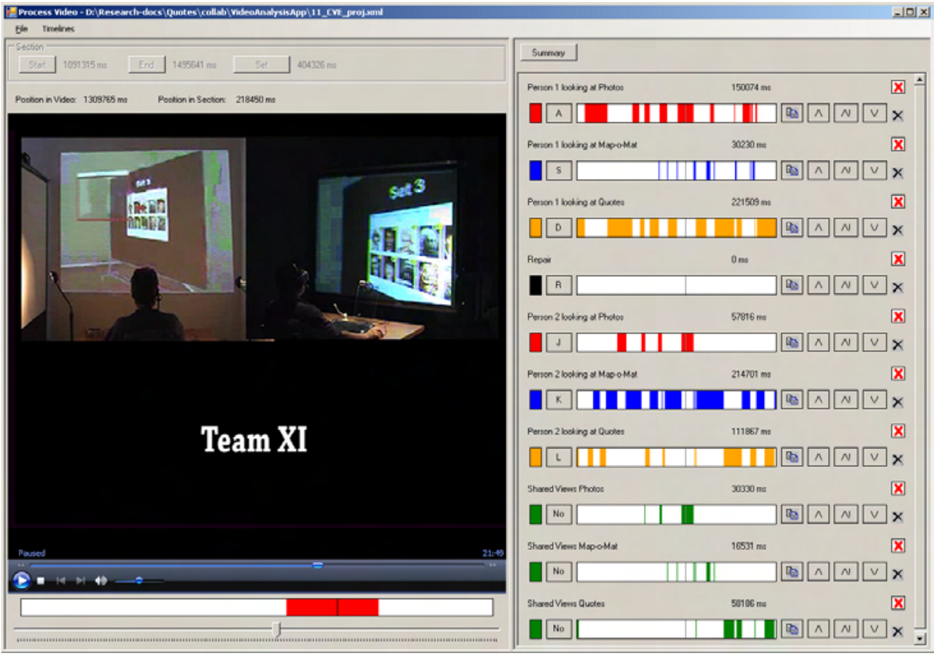


Figure 7.5: The video analysis application. The bars on each time line on the right represent different views for each of the participants that can then be compared for overlaps. There is a view overlap in the snapshot depicted, since both participants are looking at the photos in that moment.

Social presence

There was a significant main effect ($F(3,102)=39$, $p<0.001$) between the four different media. Post-hoc comparisons showed that social presence was significantly higher in FtF ($M=6.1$, $SD=0.7$, $p<0.001$) than sVC ($M=4.4$, $SD=1.1$), vCVE_desk ($M=4.3$, $SD=1.0$), and vCVE_im ($M=4.4$, $SD=0.9$). However, none of the mediated conditions showed significant differences in pair-wise comparisons.

There was also a significant interaction Medium×Gender, $F(3,102)=5.8$, $p=0.001$, according to which social presence of a medium decreased with its remoteness for female participants, while it increased for the male participants (see Figure 7.6).

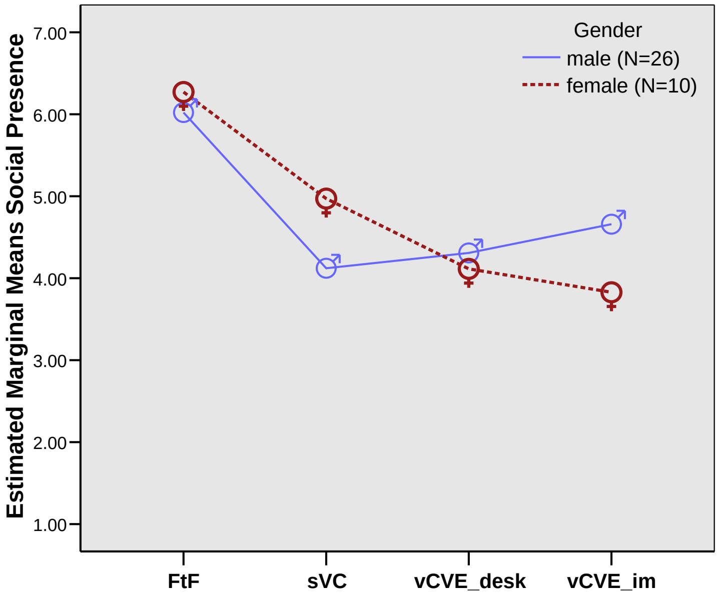


Figure 7.6: Social presence by gender and medium.

Physical presence

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

There was also a significant interaction Medium×Gender, F(1,34)=4.6, p=0.039, according to which the increase of physical presence in the more immersive vCVE condition was higher for male participants (see Figure 7.7).

In the same way as was done in Section 5.3.3, the difference scores of both physical and social presence were graphically combined in a scatter plot to explore their correlation. Figure 7.8 shows the plot.

The general distribution of data points is scattered across the four quadrants, suggesting that there is no strong linear relationship between the two dimensions.

For eighteen participants (50%) both physical and social presence increased in the more immersive vCVE condition (quadrant I). One participant (3%) felt less physically present in the more immersive vCVE condi-

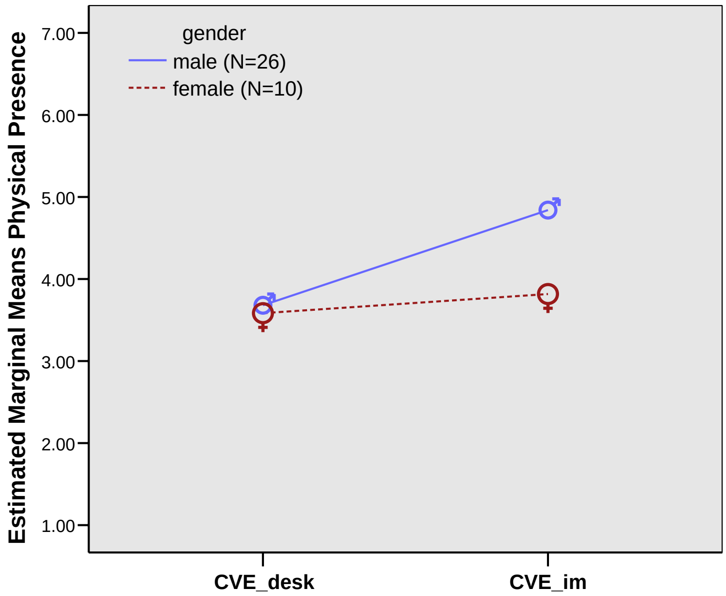


Figure 7.7: Physical presence by gender and medium.

tion, but rated social presence higher (quadrant II). Five participants (14%) experienced less physical and social presence in the more immersive vCVE condition (quadrant III). Seven participants (19%) had a stronger sense of physical presence in the more immersive vCVE, while considering it to support less social presence (quadrant IV). For five participants (14%), social presence or physical presence was independent of an increasing or decreasing sense of physical presence.

Taken together, the increase or decrease of physical presence matched the increase or decrease of social presence in 62% (data points in sector I and III) of all cases.

Copresence, awareness, and ease of use

The principle components factor analysis of the eleven items produced two factors.

The first factor emerged as a fusion of the copresence and awareness items (Q1, Q3, Q4, Q5, Q6) and accounted for 50% of the total variance. Item Q4 "I always had a good sense of what my partner was doing" loaded the highest

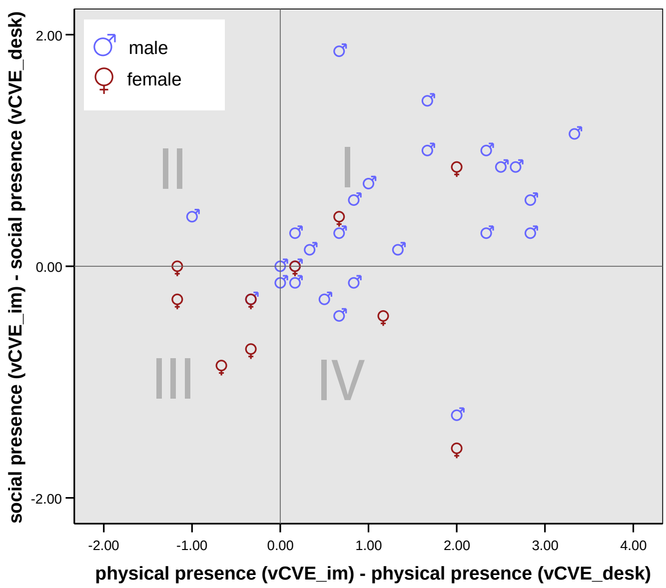


Figure 7.8: Differences in physical presence and social presence between the two vCVE conditions, marked by gender.

(0.83). The factor was therefore labelled "awareness". Internal consistency for this factor was sufficient (α=0.88).

The second factor could clearly be interpreted as "ease of use". It comprised four items (Q2, Q7, Q10, Q11) accounting for 11% of the total variance. Item Q10 "I was often confused" loaded the highest of this factor (0.75). Internal consistency for this factor was borderline (α=0.70).

Items Q8 and Q9 loaded equally on both factors and could therefore not be clearly interpreted. They were dropped from further analysis.

Awareness: a significant main effect was found across the four different media in an analysis of variance with Greenhouse-Geisser correction, F(2.4, 81.6)=55, p<0.001. Post-hoc comparisons showed that awareness was significantly higher in FtF (M=6.5, SD=0.76, p<0.001) than sVC (M=3.8, SD=1.3), vCVE_desk (M=4.0, SD=1.2), and vCVE_im (M=3.9, SD=1.3). However, none of the mediated conditions showed significant differences in pair-wise comparisons.

Furthermore, there was a significant interaction Medium×Gender, F(2.4, 82)=3.7, p=0.022, according to which the sense of awareness during collaboration decreased with the remoteness of an interface for female participants, while it increased slightly for the male participants (see Figure 7.9).

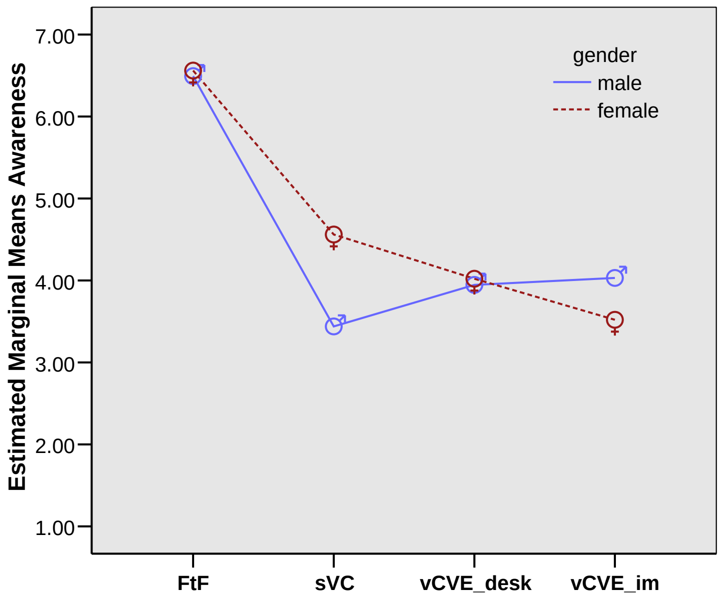


Figure 7.9: Awareness by gender and medium.

Ease of use: a significant main effect was found across the four different media, F(3,102)=17, p<0.001. Post-hoc comparisons showed that the ease of use was significantly higher in FtF (M=6.2, SD=0.76, p<0.001) than sVC (M=5.6, SD=0.81), vCVE_desk (M=5.4, SD=0.93), and vCVE_im (5.2, SD=1.3). However, none of the mediated conditions showed significant differences in pair-wise comparisons.

In the mediated conditions, both female and male participants considered the ease of use to decrease with the remoteness of the interface. This tendency was more distinct for female participants (see Figure 7.10). However, no significant interaction between Medium×Gender was found.

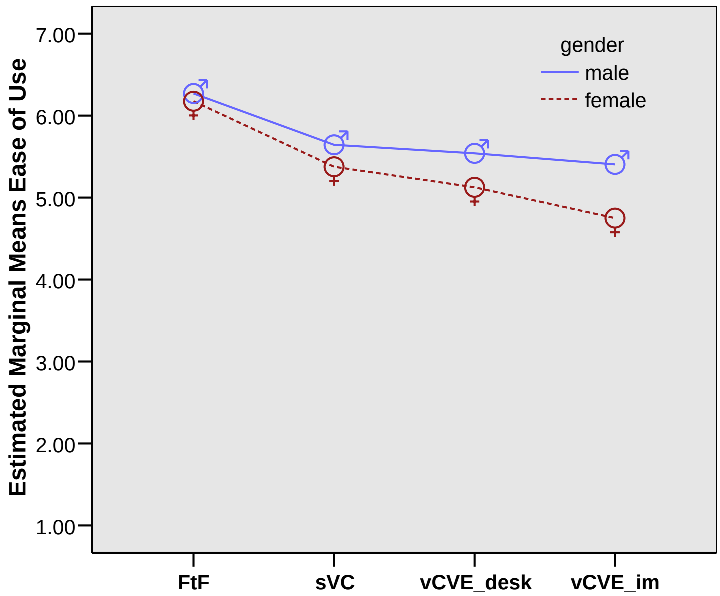


Figure 7.10: "Ease of use" by gender and medium.

Questionnaire scale validity

A principle components factor analysis extracted one single factor for social presence, explaining 61% of the variance. The uni-dimensionality of this construct was therefore confirmed. The item "active–passive" had the lowest factor loading (0.53) and was excluded from the analysis. It might have been associated with action in terms of navigation in the virtual environments rather than communication and thus did not align well with the other items targeting social presence. Internal consistency for the factor social presence based on the remaining seven items was good (α=0.91).

A principle components factor analysis extracted one single factor for physical presence, explaining 65% of the variance. The uni-dimensionality of the this construct was therefore confirmed. Internal consistency for the factor physical presence based on all six items was sufficient (α=0.89).

7.4.2 Results preference

An initial Friedman test was performed including the rankings of all conditions. Average rankings revealed that 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, χ_r^2=40.8, df=3, N=36, p<0.001).

A second Friedman test was run with only the rankings of the mediated conditions. It did not indicate a clear preference for any of the three mediated interfaces (Friedman test, χ_r^2= 2.4, df=2, N=36, p=0.30).

Two more Friedman tests were therefore carried out with the rankings of the three mediated conditions for male and female participants separately. No clear preference emerged for the male participants (Friedman test, χ_r^2=0.23, df=2, N=26, p=0.89). However, female participants clearly preferred the standard video-conferencing interface (M=1.2) more than interface vCVE_desk (M=2.0) and vCVE_im (2.8), (Friedman test, χ_r^2= 12.8, df=2, N=10, p=0.002).

Most participants (male and female) preferred the FtF condition. However, eight participants favoured one of the mediated conditions over FtF, arguing that it involved less effort to press buttons than to walk around a table.

Many participants saw the main benefit of the sVC condition in the fact that the face of the other person was visible all the time. This, however, was considered more or less important by different participants:

- One male commented for example that he hardly looked at his friend's face because he "knew him well". Other male participants confirmed that they only rarely looked at the other's face and concluded that the sVC condition only felt to them as if they were on the phone with the other person. Communication was "done by talking" only.

- In contrast, in particular some of the female participants frequently mentioned that they looked at their friend's video and considered the ability to see details of the other's face as being beneficial. One female participant stated that the sVC condition was great, because she could see her friend and talk to her as if they were on the phone.

Navigation issues were the most frequently mentioned problems experienced in the vCVE conditions. Using a space mouse was considered “hard”, “unfamiliar”, or “distracting” by many participants. However, some participants liked the feeling of presence in the virtual room, felt like they had more options for interaction, and liked the ability to move very close to the photos on the photo board.

While the more immersive vCVE setup increased the “intensity” and the “fun” of the experience for some, others explicitly stated that it did “not add anything in terms of the collaboration”. Having to wear glasses was considered another drawback by some: one female pointed out that “it is all about comfort zone and fashion”. Two participants mentioned moderate signs of motion sickness after the immersive vCVE condition and therefore gave this condition the lowest ranking.

Many of the participants also explicitly mentioned the novelty aspect of the space mouse and stereo projection which, they admitted, might have had a negative impact on their rankings.

7.4.3 Transcription analysis

Approximately 23,000 words were transcribed from the audio files of the last eight teams by the author. One of the male teams who would have been part of the last eight teams was dropped because the participants were not sufficiently fluent in English. This team was replaced with the next succeeding male team. On average, 68% of all transcribed words matched words or word stems included in the default dictionary of the lingusitic analysing tool.

The results reported in the following are divided into five sections. The first three sections present findings of the categories “standard linguistic dimensions”, “psychological processes”, and “relativity”, following the general division of categories suggested by the analysing tool LIWC2001 (see Appendix C.2). The fourth and fifth section will then cover the results obtained from the additional dictionary that was specially created for this experiment.

The transcripts were analysed for all main word categories included in the dictionary of the tool. Further sub-dimensions were investigated which were expected to show relevant effects in spoken language (based on Kramer et al.'s (2006) findings). The results obtained (except for word count) are in percentages, that is, the relative occurrence of words of a certain category in the transcripts.

Standard linguistic dimensions

The results of the standard linguistic dimensions are shown in Table 7.1. The table lists the dimensions tested and shows where significant effects or interactions occurred.

	Effects			Post-Hoc Comparisons
Dimension	Medium	Gender	M×G	Medium
WC	p<0.001	*	p=0.043	FtF<sVC, FtF<vCVEdesk, FtF<vCVEim
WPS	*	*	*	–
Overlaps	p<0.001	*	*	FtF<sVC, vCVEdesk<sVC
Qmarks	*	*	*	–
Pronouns	*	p=0.006	*	–
Self	*	*	*	–
Other	*	p<0.001	*	–

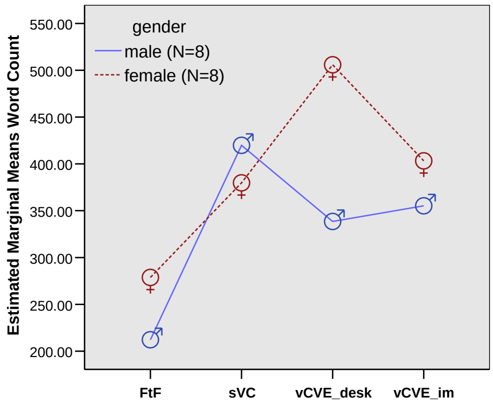
* = no significant difference

Table 7.1: Effects in category "Standard Linguistic Dimensions".

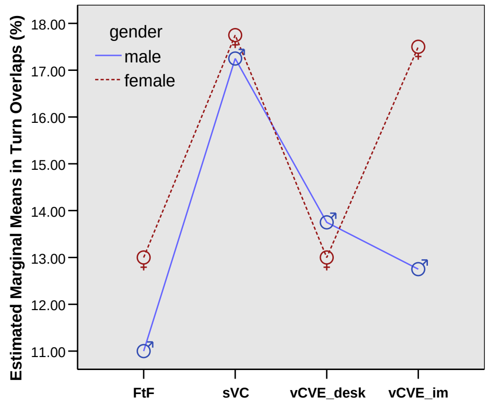
The analysis of total word counts (LIWC-category "WC", see Appendix C.2) revealed that significantly less words were spoken in the FtF condition than in the three mediated conditions. Furthermore, there was an interaction Gender×Medium (see Figure 7.11(a)). While male participants spoke most in condition sVC, female participants spoke most in condition vCVE_desk. The words per sentence (LIWC-category "WPS", see Appendix C.2) did not show any significant effects.

Turns in which both participants spoke over one another at some point also differed between the media conditions. There were significant more turn overlaps in all teams in condition sVC than in conditions FtF and vCVE_desk. Furthermore, according to Figure 7.11(b), female teams had more overlapping turns in condition vCVE_im. No significant effects were found for the relative occurrence of questions (LIWC-category: "Qmarks").

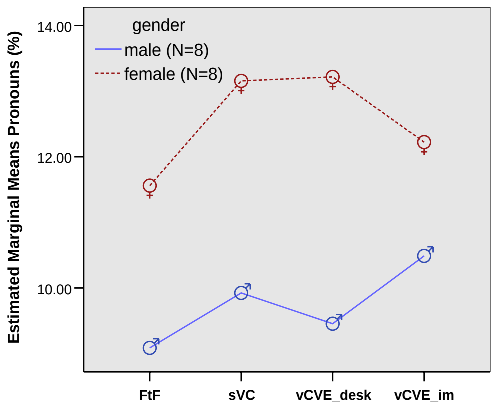
There was a significant difference of the occurrence of pronouns between genders: female participants used more pronouns (LIWC-category "Pronouns", e.g. he, our, they, you're) than males in their speech (see Figure 7.11(c)). This between-gender difference was confirmed for third person pronouns (LIWC-category "Other", e.g. she, their, them), (see Figure 7.11(d)), whereas no differences could be found for first person pronouns ("Self", e.g. I, we, me).



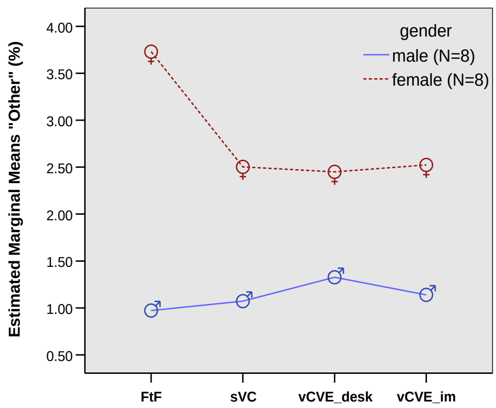
(a) Word count (WC)



(b) Turn overlaps



(c) Use of pronouns



(d) References to "others"

Figure 7.11: The standard linguistic dimensions that showed significant effects.

Psychological processes

The results obtained based on word occurrences that fall into the categories "Psychological Processes" are shown in Table 7.3. The table contains the dimensions tested and shows where significant effects or interactions occurred.

Dimension	Effects: Medium	Effects: Gender	Effects: M×G	Post-Hoc Comparisons: Medium
Affect	p=0.012	*	*	–
Cogmech	*	*	*	–
Senses	*	*	*	–
Social	*	p=0.01	*	–

* = no significant difference

Table 7.2: Effects in the category "Psychological Processes".

There was a significant effect of the medium used to collaborate on the relative use of words belonging to affective or emotional processes (LIWC-category "Affect", e.g. happy, ugly, bitter; see Appendix C.2). As Figure 7.12(a) shows, people in the FtF condition used affective words less frequently than in the mediated conditions. However, these differences did not reach significance in post-hoc comparisons.

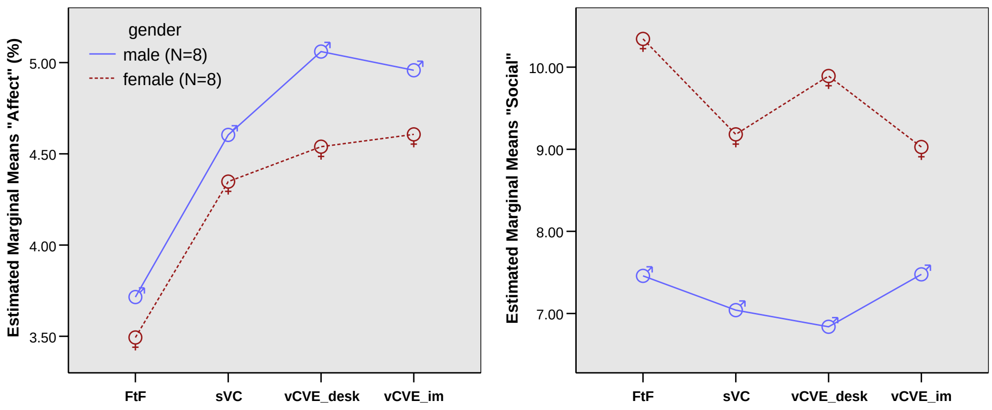
Words from the categories cognitive processes (LIWC-category "Cogmech", e.g. cause, know, ought), or sensory and perceptual processes (LIWC-category "Senses", e.g. see, touch, listen) did not show any effects.

Female participants used words of the category social processes (LIWC-category "Social", e.g. talk, us, friend) more often than male participants (see Figure 7.12(b)).

Relativity

The two LIWC-dimensions "Space" and "Motion" were tested for effects in the category "Relativity". Table 7.3 shows where significant effects or interactions occurred.

There was a significant effect of the relative occurrence of words belonging to the word dimension "Space" (e.g. around, over, up) across the four conditions. Post-hoc comparisons showed that participants used words of



(a) Affective or Emotional Processes (Affect)

(b) Social Processes (Social)

Figure 7.12: Dimensions in the category psychological processes that showed significant effects by gender and medium.

Dimension	Effects: Medium	Effects: Gender	Effects: M×G	Post-Hoc Comparisons: Medium
Space	p<0.001	*	*	vCVEim>FtF, vCVEim>sVC, vCVEdesk>FtF, vCVEdesk>sVC
Motion	p=0.033	*	*	–

* = no significant difference

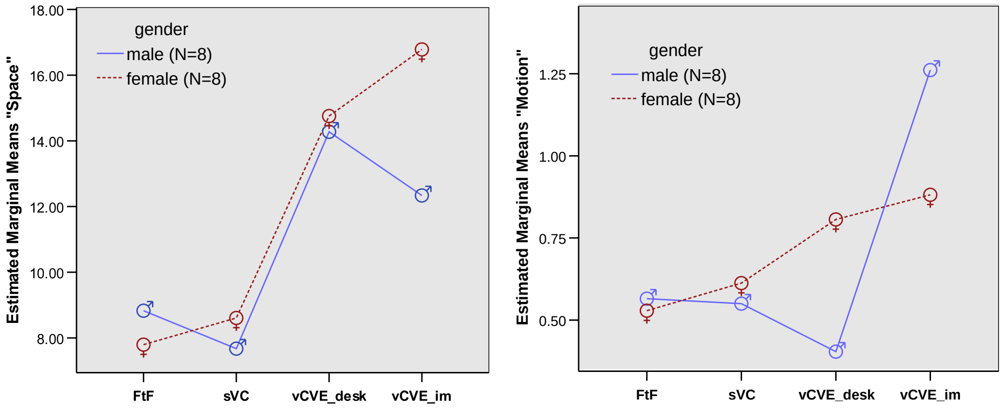
Table 7.3: Effects in category "Relativity".

that category more often in both vCVE conditions than in conditions sVC and FtF (see Figure 7.13(a)).

There was also a significant effect of word occurrences of the dimension "Motion" (e.g. walk, move, go) across the four conditions. Figure 7.13(b) suggests that words belonging to this category were used more often in condition vCVE_im. However, post-hoc pair-wise comparisons did not yield any significant differences.

Other dimensions

Three additional word categories were defined in a dictionary that was created specially for this experiment. These categories were laughter (e.g.



(a) Space (b) Motion

Figure 7.13: Dimensions in the category "Relativity" that showed significant effects by gender and medium.

"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. Table 7.3 shows the results.

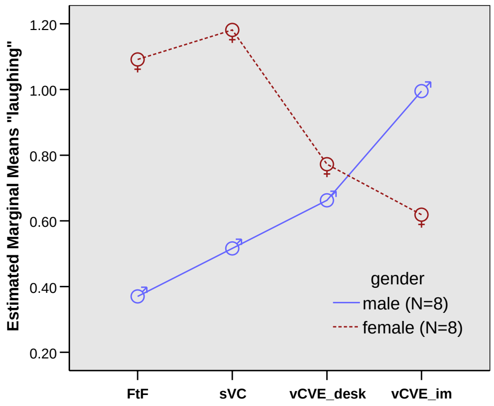
	Effects			Post-Hoc Comparisons
Dimension	Medium	Gender	M×G	Medium
laughter	*	*	p=0.01	–
local deixis	p=0.001	*	*	FtF>sVC, FtF>vCVEim
remote deixis	*	*	*	–

* = no significant difference

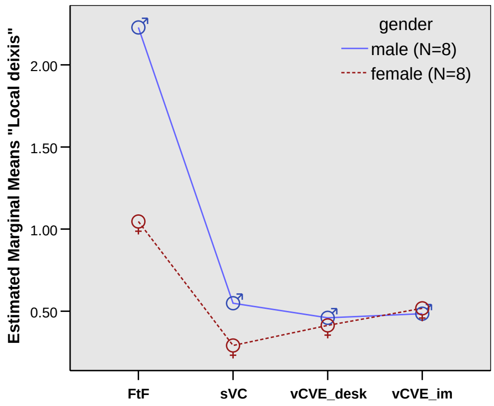
Table 7.4: Effects in additional word categories which were defined in a separate dictionary.

There was a significant interaction between gender and media 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 (see Figure 7.14(a)).

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. According to Figure 7.14(b), in the FtF condition, male participants used local deixis more often than female participants. No effect was found with regards to the occurrence of remote deixis.



(a) Laughter



(b) Local deixis

Figure 7.14: Additional linguistic dimensions that showed significant effects by gender and medium.

Referencing of quotes

Participants verbally referenced a quote in three different ways.

1. They read out the whole quote to the other participant. This is verbally expensive, but does eliminate uncertainty and does not assume a certain level of common ground.

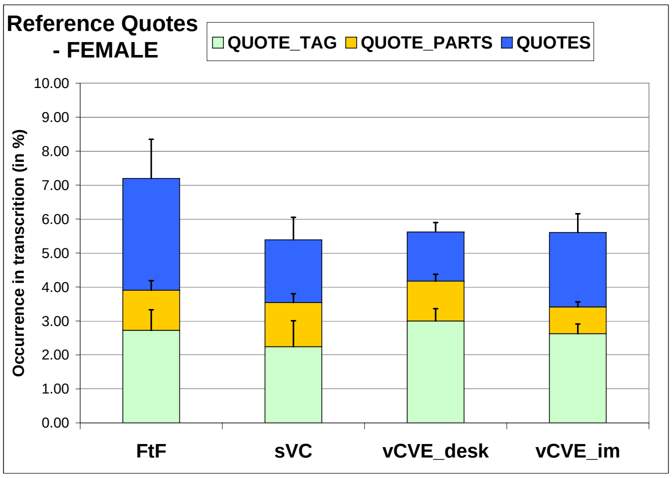
2. They referred to a quote by reading or repeating only a significant part of it. This was verbally cheaper, but assumed that the other participant was already familiar with the quote and would thus identify it. This way of referencing therefore requires a higher level of common ground.

3. They referred to a quote only by the tag that was displayed on the side of each quote. While this way of referring to a quote involved the least verbal effort, it required a high level of common ground and interpersonal awareness, since a tag could only be understood by the other person if he or she looked at the quotes in that moment.

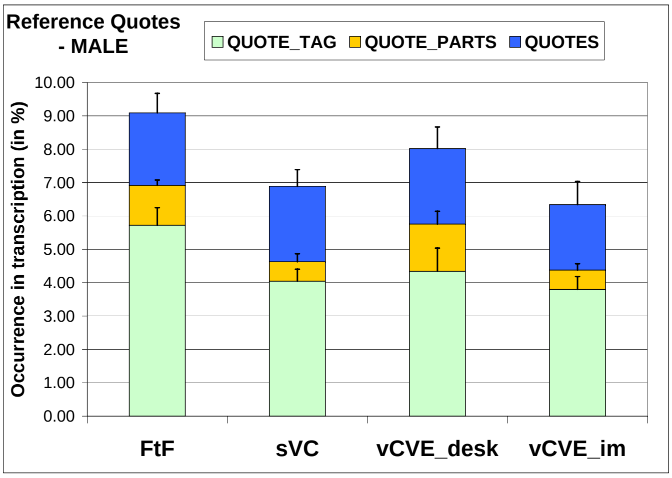
The relative occurrences of each of these three ways of referencing were determined, averaged by condition, and are displayed accumulatively together with standard error bars in Figures 7.15(a) and 7.15(b), showing two separate charts for female and male participants.

The relative occurrence of all types of referencing together is highest in FtF for both male and female teams (within-media effect, $F(1,14)=6.4$, $p=0.024$).

Moreover, 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.



(a) female references to quotes



(b) male references to quotes

Figure 7.15: 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).

7.4.4 Video analysis

The averages of the view overlaps and standard errors based on four male and four female teams are shown in Figure 7.16. No significant main effect or interactions were found. However, there was a significant between-subjects effect (F=18, p=0.006) between male and female teams in condition FtF.

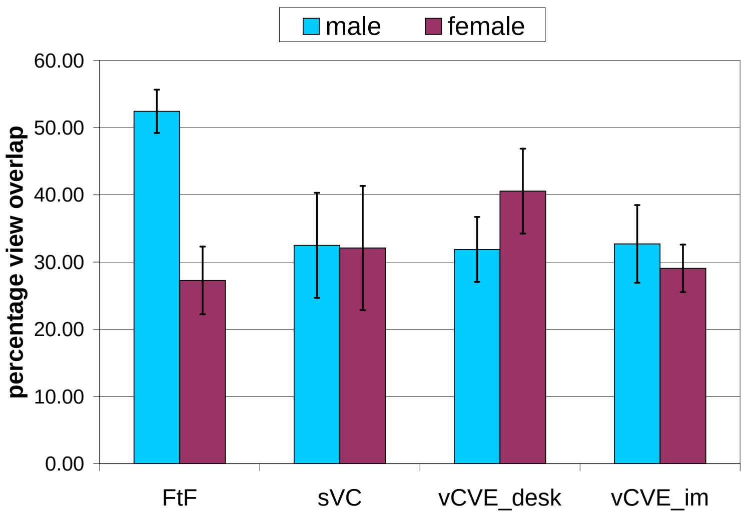


Figure 7.16: Average of relative view overlap and standard error by gender and medium.

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 and being talking about what was lying in front of them (see Figure 7.17(a)).

Female participants, on the contrary, often positioned themselves at opposite ends of the tables where one had access to the photos, the other one to the quotes, and then talked about it while facing each other (see Figure 7.17(b)).

(a) Male friends collaborating in a "side-by-side" arrangement

(b) Female friends collaborating in a "face-to-face" arrangement

Figure 7.17: Typical positioning for female and male team members in condition FtF.

7.5 Discussion

The findings suggest a complex interplay of several contributing factors that have played a role in this experiment. The following presents an attempt of a comprehensive interpretation.

7.5.1 The impact of the "remoteness" of a video-CVE

The key question of this experiment was if more "transporting" video-CVEs induce a higher sense of physical presence and afford a user experience and collaborative behaviour that is closer to being face-to-face. This section therefore focuses on the direct comparison of effects found between conditions vCVE_desk and vCVE_im.

Participants reported a higher sense of physical presence in the more remote condition vCVE_im, which confirmed the expected impact of transportation of an interface on a sense of "being there" at the remote space.

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. A potential cause of the gender effect is discussed in the following section.

When looking at the correlation between physical presence and social presence difference scores, a strong linear correlation like the one obtained for friends in the experiment "Dream house" could not be replicated. Instead, the tendency of physical presence and social presence matched in only 62% of all cases. The correlation was negatively affected through the introduction of new side effects. For example, two participants reported that they experienced light forms of motion sickness while feeling present in the immersive vCVE. This lead to discomfort and had in turn a negative impact on their social presence ratings of that medium.

The technical overhead of condition vCVE_im sometimes also impaired collaboration. Participants (especially female participants) did not like to wear the shutter glasses, others were not used to the large field-of-view display, resulting in lower ratings for the "Ease of Use" of that condition.

The analysis of word counts could not replicate the presence correlations with the word categories found by Kramer et al. (2006), 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 "remote" 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.

7.5.2 Gender effects

Results obtained in this experiment substantiated the presumption of experiment "Dogs & Owners" that gender differences in communication impact perception of communications media. This section dicusses possible explanations based on the data gathered in this study.

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 (1982) characterisation 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."

In consideration of these given differences it can furthermore be explained that the different collaborative systems used in this study supported these collaborative styles more or less well, leading to the observed gender differences and interactions in the mediated conditions.

The sVC condition allowed participants to see their partner's faces at all times, but did not provide visual awareness cues. This supported the interpersonal, verbal collaboration style adopted by female teams, who therefore rated both social presence and awareness higher in this 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 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. (1968), who studied the effects of visibility on interaction in a dyad. They encountered "considerable sex differences" according to which females were less comfortable in situations where they could not see their counterparts.

7.5.3 Transcription analysis

The transcription analysis did not confirm Kramer et al.'s (2006) findings. However, other interesting differences emerged between non-mediated and mediated communication. Communication in the FtF condition involved fewer spoken words and contained fewer words belonging to affective or emotional processes, suggesting that the mediated conditions created a verbal overhead for coordinating collaboration and expressing emotion due to the absence of non-verbal forms of communication. In the mediated conditions, male participants spoke most words in condition sVC, suggesting this condition to be the least efficient with regards to verbal effort. In contrast, female participants spoke most words in condition vCVE_desk.

The highest occurrence of turn overlaps in the sVC condition indicates a more "chaotic" communication style in this condition, which is, however, normally found in face-to-face talk. Conversations in conditions FtF and vCVE_desk therefore have been more "formal", which partly contradicts commonly reported results of speech comparisons between FtF and mediated communication. The reasons for this may lie in the slightly unusual situation of talking to each other while moving around a table.

Other differences emerged between female and male teams that are harder to interpret. For example, female teams used more pronouns, especially references to others such as "she", "their", "them", while male teams used more references to quotes (see Figure 7.15(b)). One possible explanation could be that, during their conversations, female participants tended to pick a quote and discuss which of the celebrities might match it, whereas male participants tended to pick a celebrity and discuss which of the quotes he or she could have said. Consequently, more third person references to celebrities were found in female conversations, while a higher occurrence of references to quotes was found in conversations of male participants.

7.5.4 Reciprocal awareness issues in the vCVE conditions

One problem that surfaced in the vCVE conditions was that the visibility of the other avatar was based on chance encounters. Participants focused on their primary task and navigated between the billboards where they then often met halfway, but did not make a conscious effort to look for the other participant.

Consequently, they did not rely on their partner being able to see what they were doing and therefore often narrated their own actions within the environment like “I am back at the quotes board”, or “I am walking over to the Map-o-Mat”, which may have lead to the high occurrence of words of the categories “Motion” and “Space” in the vCVE conditions.

However, in situations when their partner actually looked at their avatar, these verbal narrations were needless verbal effort. The problem here was the lack of reciprocal awareness, that is, *participants were not aware that their partners were aware of their actions.*

Based on the same problem, verbal effort was also 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 Eastwood 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.

In conclusion, 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 his 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 can not provide the same level of peripheral awareness mainly because of a limited field of view and the lack of other sensory stimuli such as smell or touch.

7.6 Experiment summary

This chapter presented the results of a study comparing a desktop video-CVE with a more immersive stereo-projected video-CVE, a standard video-conferencing system, and a face-to-face condition. Participating teams consisted of two same-gender friends.

Already in the face-to-face condition, a considerable difference in the collaborative style between male and female teams emerged which may explain further observed differences in user experience aspects of the mediated conditions. Of the mediated conditions, male participants rated the immersive video-CVE as supporting the highest level of social presence and awareness, followed by the desktop based system and the standard video-conferencing tool. This suggested that a higher level of transportation leads to a more face-to-face-like experience. This, however, was not the case for female participants, who rated these systems in exactly the opposite order in social presence, awareness, and preference.

Collaborative behaviour, assessed by analysis of communication patterns and view coordination, highlighted interesting differences between mediated and unmediated conditions, and between male and female teams. However, it did not provide clear evidence concerning which of the conditions supported collaboration which was closest to being face-to-face.

This study concluded the experimental part of this thesis. The next chapter will discuss the overall findings and will propose directions for future work.

Chapter 8

Discussion

The overarching goal of this thesis is to gain understanding of human factors of video-CVEs by investigating if geographically dispersed people who collaborate in video-CVEs feel and behave more similarly to being face-to-face than if they collaborate through standard video-conferencing systems. The four user experiments described in the previous four chapters assessed and compared aspects of user experience and collaborative behaviour as afforded by several VMC interfaces in order to answer the research questions raised.

This chapter discusses and draws conclusions from the main findings of the studies and proposes possible directions for future work. First, Section 8.1 summarises the findings that indicated if spatial virtual interaction afforded by video-CVEs *feels* more similar to being face-to-face than normal video-collaboration. Section 8.2 then discusses findings which indicated whether participants collaborating in video-CVEs *behave* in a more similar way to being face-to-face than when working together through a normal video-conferencing connection. Section 8.3 introduces and illustrates the fundamental trade-off between *spatiality* and the ease of interaction within a tele-collaboration interface which surfaced in the experiments. It also proposes that the value of video-CVEs can be conceived of in terms of a cost-benefit ratio, with the costs and benefits being subjectively experienced by different user groups. In light of this understanding of the value of spatial virtual interaction, Section 8.4 points out directions researchers should pursue next. Finally, Section 8.5 comments on limitations of the research approach and the measurements applied.

8.1 User experience of spatial virtual video-conferencing

One component of the research hypothesis was to determine whether video-CVEs create a collaborative *experience* that is closer to the feeling of being face-to-face than regular video-conferencing interfaces.

To study this research question, several subjective rating scales of the user experience of tele-collaboration were first trialled for their applicability in comparisons of video-CVEs and standard video-conferencing tools (see Chapter 4). One subjective scale that was found to be particularly sensitive was Short et al.'s (1976) measure of social presence. This assesses the "salience of the other in the interaction" based on a participants' subjective ratings of the medium itself (see Section 2.2.1). Social presence was thus measured repeatedly as one of the core user experience dimensions in all four studies.

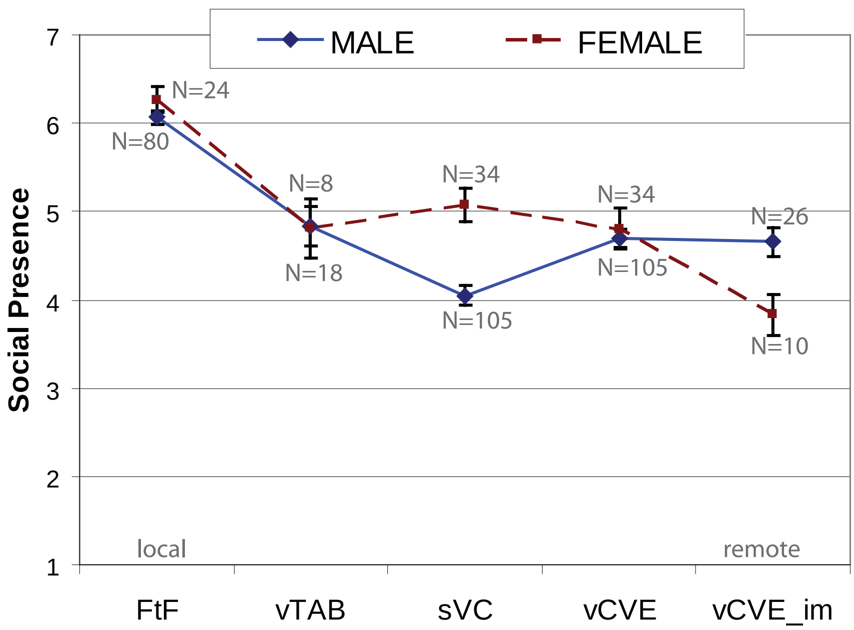


Figure 8.1: Social presence, averaged over all experiments.

Figure 8.1 integrates results obtained from all studies into one chart by averaging all social presence scores collected for each condition. Since some of the conditions were repeated in the four experiments, the number of data points which contributed to each average score differs considerably across conditions and is thus displayed in the chart. Averaging social presence ratings across experiments does not take into account the differences of the instantiations of the interfaces, the mixed and same-gender group constellations, and the different natures of the experiment tasks. However, the standard errors of the averages of FtF, sVC, and vCVE are small compared to the effects that emerge between the conditions. Therefore, it can be argued that the differences that can be observed between conditions can be ascribed to more substantial, higher-level differences based on the nature of the interface types.

8.1.1 Gender effect

One salient finding taking all experiments[1] into consideration was that social presence of VMC interfaces were subject to a gender difference. As previously discussed in Section 7.5.2 this may be the result of variations of gender specific communication styles which are more or less supported by the individual interfaces. The differences encountered suggested that firstly, for women, standard video-conferencing interfaces yield more social presence than for men, and second, that the addition of spatiality entails a reduction of social presence for women, while it promotes more social presence for men.

Compared to men, women use more facial expressions (Hall 1990, page 71) and gaze at each other more, especially in same-sex dyads (Hall 1990, 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.

Gender effects have not been reported and discussed much in other cross-media studies, perhaps because not many experiments study same-gender friends, which may have fostered the occurrence of the gender effect in the

[1] Note: a significant interaction Medium×Gender was previously only detected in experiments "Dogs & Owners" and "Celebrity Quotes". The interaction Medium×Gender was not investigated in experiment "DSG" due to the small group of female participants, and did not show significant effects in experiment "Dream House".

studies here. One exception is a study reported by de Greef and IJsselsteijn (2001), who found a gender interaction according to which a talking head video conveyed more social presence for female participants than for male participants. As has been argued in this thesis, de Greef and IJsselsteijn also suspect 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", but do not further investigate the causes in more detail.

One implication of this research for the CSCW community is that future studies investigating video-mediated communication should pay more attention to gender-specific differences.

8.1.2 Spatiality and the feeling of being face-to-face

One research question asked was if spatiality correlates with a feeling of being face-to-face, as measured by social presence, copresence, and awareness. Because of the gender effect, this question must be discussed for male and female participants separately.

Collaborating through spatial video-conferencing interfaces allowed participants to use additional non-verbal communication cues, while at the same time compromising the visibility of the other partner's video and introducing the need for avatar navigation and spatial object manipulation. Findings suggest that, especially for male participants, the addition of spatiality was perceived to create a collaborative context closer to being face-to-face, suggesting that the collaborative style typically adopted by men depends on and utilises spatial properties more.

For female participants, however, the non-verbal communication cues that were made available through spatiality were perceived to be more distracting than helpful or natural, which caused the collaborative context created by spatial VMC interfaces to feel more detached from being face-to-face.

8.1.3 Transportation and the feeling of being face-to-face

Two experiments investigated the correlation between physical presence and social presence to explore whether a higher level of transportation of an interface can foster a user experience which is closer to being face-to-face. Findings suggest that the answer to this question must distinguish between teams of friends and teams of strangers, as well as between teams consisting of female participants and teams of male participants.

The first experiment (Chapter 5) concerned the differences in transportation between a standard video-conferencing interface and a desktop-based video-CVE. Findings suggest that there is a weak positive relationship between physical presence and social presence, which is more distinct when only considering teams who know each other well.

The second experiment (Chapter 7) concerned the differences in transportation between two video-CVEs which induced different levels of physical presence. Findings reveal a gender effect, in which male participants rate social presence and awareness in the video-CVE with the higher level of transportation as being closer to being face-to-face, while for female participants the higher level of transportation generates a higher level of physical presence, but at the same time distraction and a feeling of discomfort. Consequently the user experience departs further from a sense of being face-to-face.

The results of both experiments show that the value of using immersive technology for creating a face-to-face-like user experience in remote collaboration is very susceptible to the comfort people feel in an immersive experience. These comfort levels vary depending on confounding factors like motion sickness, technology aversion, or because it may create a user experience that is too intimate for some, especially when meeting a person one does not know well.

8.2 User behaviour in spatial virtual video-conferencing

Another component of this research was to determine if people who collaborate in video-CVEs *behave* more similarly to being face-to-face than with regular video-conferencing interfaces.

This hypothesis was not supported. The majority of speech patterns assessed in the experiments "Dogs & Owners" and "Celebrity Quotes" either did not find any differences between communication patterns in the mediated conditions or saw a higher resemblance between standard video-conferencing communication and face-to-face. There were weak indications that added spatiality of an interface may have helped male teams to reduce their number of spoken words and turn overlaps (see experiment "Celebrity Quotes"). Other positive impacts of spatiality or transportation on communication patterns were either undetected with the measurements used, or were overshadowed by verbal clutter caused by usability issues.

The fact that the task efficiency and collaborative performance were not improved in video-CVEs, despite a higher level of gaze awareness, provides an important lesson for the CSCW community. This is that research trying to improve VMC should not only focus on gaze and gaze detection, but should also attribute equal importance to improving usability, gesture support, and other forms of body-language.

8.3 Striving for the gold standard: a trade-off

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

However, when striving 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.

Such usability issues were also previously identified as one of the major shortcomings and distractions of CVEs. In the words of McGrath and Prinz (2001) "navigating 3D spaces can be compared to talking while trying to reverse a car – we *can* do it but we have to concentrate on the navigation so hard that our speech slows down." Interaction problems concerning navigation, and picking and selecting objects were furthermore mentioned by Normand et al. (1999). Greenhalgh and Benford (1995) acknowledged that navigation difficulties combined with a limited view hampered the ability to use gaze direction to negotiate turn-taking conversations.

Figure 8.2 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.

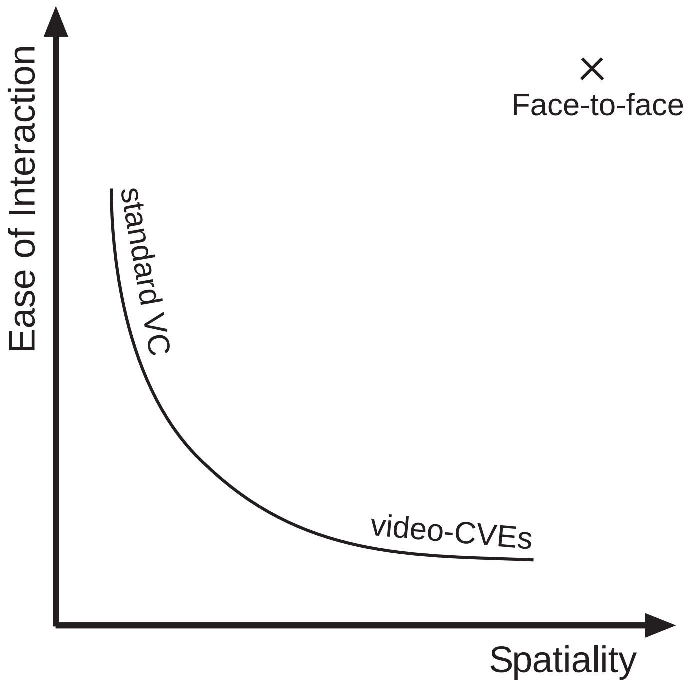


Figure 8.2: The trade-off between supporting either the spatiality or the ease of interaction of face-to-face collaboration.

Many participants in the experiments preferred the spatial video-CVE over the easier-to-use standard video-conferencing condition (e.g. 20 out of 28 participants preferred the video-CVE over a standard interface in experiment “Dream House”). This suggests that for them, the additional cognitive effort that the video-CVE demanded from its users was outweighed by the benefits it delivered in return. The value of spatial virtual video-conferencing should therefore not only be assessed merely by its usability, but rather in terms of a cost-benefit ratio which also takes advantages that can be gained into account. Figure 8.3 shows a matrix representing this concept.

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.

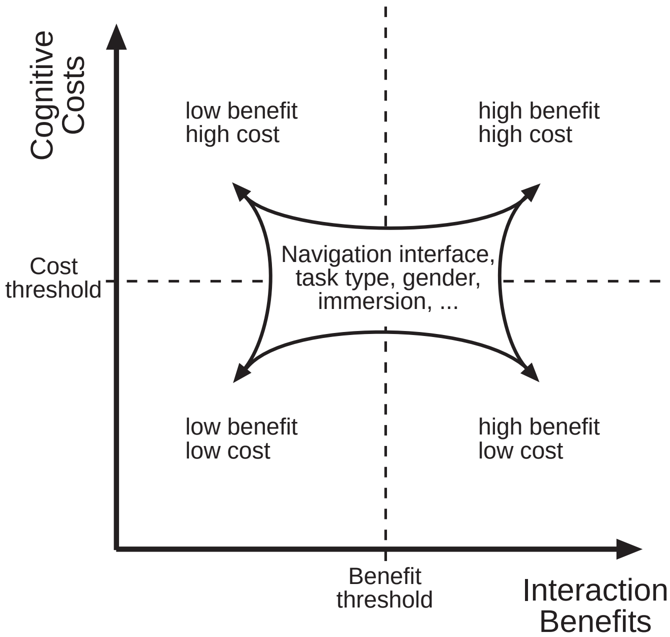


Figure 8.3: The cost-benefit matrix for spatial tele-collaboration.

- 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 experiment "Celebrity Quotes", the benefits of spatial virtual interaction support the collaborative requirements of men more than those of women and are therefore appreciated more by men.

In the experiments conducted, this 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 – these are worth investigating in future work.

8.4 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.

8.4.1 Possible directions for human factors research

Research into video-CVEs is arguably still in its infancy, and this thesis is only one of the first of many steps that need to be followed until we fully understand the full potential of this medium. There are several paths researchers could follow that would extend the findings presented in this work.

Exploring further demographic factors

The gender-effects observed in the studies in this work 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.

Spatial audio

Spatial audio was not focus of this research. However, spatial audio is likely to have an impact on social presence. The extent to which spatial audio adds to social presence in video-CVEs is an empirical question and should be investigated in future user studies.

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.

Field study

All the user data gathered in this thesis was collected in laboratory studies. The advantage of laboratory research is the high degree of control it gives to the experimenter. However, user tests that take place in a fully controlled laboratory environment may not always resemble situations we encounter in daily living and results therefore may not generalise as easily.

To complement and extend the findings of this thesis, field studies into the value of video-CVEs for everyday communication and collaboration should therefore be conducted over several weeks involving real remote collaboration projects.

Maturation of measures and instruments

While this thesis mainly relied on subjective responses and video analysis, there are other measures, including more elaborate interviewing techniques, or the assessment of physiological responses that could offer new and interesting insights.

Research advances into understanding the value of spatial virtual interaction depends on reliable and sensitive measures. New instruments and measures should therefore be constantly developed and existing measures, like the Networked Minds Measure of social presence (Biocca et al. 2001), further improved.

8.4.2 Possible directions for video-CVE interface design

The findings of this research also revealed some usability bottlenecks and awareness issues which should be addressed in order to improve the value of video-CVEs for remote collaboration.

Improving navigation

Navigating through a virtual environment by means of a computer mouse or space mouse poses the biggest additional cognitive effort for users of video-CVEs. Researchers should therefore investigate easier, more natural navigation techniques that decrease this additional work load. In this respect, head tracking techniques like applied by Vertegaal (1999), or the one deployed in experiment "Dogs and Owners", present promising alternatives to mouse-based navigation.

Although still in its infancy, research into brain-computer interfaces demonstrated that participants were able to navigate through a virtual environment simply and exclusively "by thinking" (Leeb et al. 2006, Friedman et al. 2007). Future advances in that area could provide interesting alternatives also for a more intuitive navigation for video-CVEs.

Improving object manipulation

Primitive mouse-based object manipulation does not support the same simultaneous, two-handed, and lightweight interaction people take for granted when collaborating in face-to-face situations.

Therefore, novel spatial interaction mechanisms must be explored that can reduce the mental effort and keep up with the highly interactive nature of face-to-face-like communication. In this respect the tangible user interfaces (TUI) (e.g. Ishii and Ullmer 1997, Rekimoto et al. 2001, Waldner et al. 2006) offer promising alternatives that the author explored in previous research (Hauber et al. 2004).

Investigating cues for reciprocal awareness

The restricted peripheral awareness of the CVEs used in the experiments prevented users from noticing other avatars that were 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.

Other approaches which may also enhance reciprocal awareness in video-CVEs were proposed by Fraser et al. (1999). They included superimposed visualisation of the field of view of others to disambiguate the inference of the gaze of others, and used graphical peripheral lenses to enhance peripheral vision.

Combining standard video-conferencing interfaces with video-CVEs

The feasibility and consistency constraints of a new generation of hybrid VMC interfaces could be explored that allows video-CVEs to connect to standard video-conferencing interfaces. Such systems could allow remote users to individually choose their preferred interface for a given situation.

Artefacts of a shared workspace in a standard video-conferencing system could for example automatically be spatially arranged in a video-CVE and the navigation of a person using the standard interface could be directly mapped to move his or her avatar onto his video-avatar, who would then be moving between the spatially arranged artefacts, visible for others who choose to be in the video-CVE visualisation mode. Participants could swap between interface representations on the fly, and so adapt the user experience to their actual collaborative needs and personal preferences.

8.5 Critical reflection

This section critically reviews the measures used and the circumstances they were applied in, and discusses possible threats to the *reliability* and *validity* of the results and their interpretation.

A psychological measure is *reliable* if its results are repeatable when the behaviours are remeasured (Goodwin 1995, page 96). Reliability is important, because it enables one to have some confidence that the measure taken is close to the "true" measure.

A psychological measure is *valid* if it measures what it has been designed to measure (Goodwin 1995, page 96). In this context, *construct validity* concerns whether the construct being measured by a particular tool is a valid construct and whether the particular tool is the best one for measuring the construct.

The concept of validity also applies in a more general context: psychological research is *valid* if it provides the understanding about behaviour that it is supposed to provide (Goodwin 1995, page 142). Therefore, also the external and internal validity of the experimental findings will be considered in the following sections.

External validity refers to the degree to which research findings generalise beyond the specific context of the experiment being conducted.

Internal validity refers to the degree to which an experiment is methodologically sound and confound free.

8.5.1 Reliability

Were the measures applied in each of the experiments consistent from one session to the other?

Several precautions were taken in accordance with best practice guidelines in experimental psychology in order to reach high reliability between the sessions of each experiment. Instructions for participants were, for example, written down wherever possible so that they would be identical for all participants. Experimenters who demonstrated the usage of an interface also followed a predefined script to make sure that same aspects were emphasised to all of the participants.

Both the order of conditions as well as the materials used in each round were balanced across participants. One potential problem for reliability, which could not be completely controlled, was differences in collaborative behaviour which might have been specific to each of the material sets that were used. Although the sets used were carefully composed to be as similar in character and challenge as possible, users might have responded slightly different to certain photos used in the experiment "Dogs & Owners", and likewise to certain quotes in the experiment "Celebrity Quotes". These differences, however, can be expected to be negligible compared to the differences in collaboration shaped by the different media conditions.

Learning effects, another potential threat to the reliability of any within-subjects experiment, were of minor concern in the studies conducted. Firstly, the primary focus of the investigations was on the process of collaboration and not on its outcome. Secondly, the type of tasks used, which involved primarily ambiguity resolution through communication, was not prone to learning effects.

8.5.2 Construct validity

Did the measures applied really measure what they were supposed to measure?

The applied social presence and physical presence scales were adapted from existing literature and were controlled and sufficiently tested for their dimensionality and reliability by carrying out confirmatory factor analyses and testing internal consistencies of every data set. It can therefore be assumed that these scales indeed measured social presence and physical presence following the definitions that underlie the measurement approaches.

Other factors such as awareness, ease of use, or copresence emerged from exploratory factor analysis and were interpreted and labelled in accordance to best practice for factor extraction, following the item with the highest factor loading. Construct validity can therefore also be assumed for these constructs.

Construct validity for the video analysis is harder to assess, especially for the measures of verbal effort based on word-counts in experiment "Celebrity Quotes", since it was the first analysis of this type which was conducted in the given research context.

8.5.3 External validity

Can the results found in the controlled laboratory environment generalise to other participants, other tasks, other environments, or other times?

Participants: Participants in the study were recruited among University staff and students from different departments ranging from engineering to forestry. The population tested in the user tests therefore therefore covered a broad spectrum of experiences and personalities which allow the results to be generalised to a fairly broad group of real-world users.

One advantage for real-world users could be that they have more time learn the systems and may become experts using them for their collaborative tasks.

One open question is, whether the results found for groups of two or three would also generalise to groups of five or more members. In bigger groups it is harder to coordinate joint action.

Tasks: The tasks used in the experiments were artificial. However, the elements of the experiment tasks may be found more generally in everyday collaboration: small group conversations, review of a virtual object, review of a set of photos, and data retrieval from different locations and operation of a shared application. The tasks chosen therefore cover a good range of activities that people engage in on a daily basis, and thus can be generalised to real-world collaboration.

Environments: The experiments "Dogs & Owners" and "Celebrity Quotes" were explicitly designed to investigate human factors with regards to the two design dimensions spatiality and transportation which are also relevant to the design of other telecommunication environments.

Based on the general consistency of the results obtained for two substantially different spatial conditions in the study "Dogs & Owners", it can be assumed that these results also generalise to other spatial environments.

The results obtained for the "remote" condition in experiment "Celebrity Quotes" have less external validity, since results obtained have to be partly ascribed not only to the remote character of the interface, but also to its specific technical implementation. Other environments, such as fully immersive CAVE environment which track participants' positions and orientation, may be capable of inducing a stronger sense of presence while being easier to navigate through. Collaboration and user experience afforded by these remote environments could then be expected to differ substantially.

Times: Would the same results show in five, ten, or fifty years time? One threat to generalising users' responses of any type of new media they are exposed to is the novelty effect that plays a role. As IJsselsteijn (2003) points out in a review of the introductions of age-old media technologies, peoples first responses to new and more realistic media such as the first photograph, first movie, or the first virtual environment have always been characterised as being very exciting, emotional, and intriguing.

Therefore, it could be expected that, as time goes by and similar types of systems such as online 3D worlds get more widespread and accepted, a "wow-effect" that may have played a role in the experiments conducted in this thesis will wear off and people will get more acquainted and comfortable operating these type of systems. Research results based on user responses toward these type of systems are therefore also likely to change over time.

8.5.4 Internal validity

Were experiments conducted methodologically sound and were confounding factors minimised?

Data analysis: The questionnaire data was collected and analysed in close accordance to standard procedures and tests and are therefore, to the author's understanding, methodologically sound.

The video analyses were conducted by the author only. This could be considered a potential problem for the internal validity of the results, since observations made are potentially subject to experimenter bias. To avoid biased results, standard procedure recommends to compare the results of at at least two experimenters who perform the same analysis separately. Since the video analyses were very time consuming, unfortunately no second experimenter was found who was willing to do the same analyses. However, when looking at the results of the video analyses in a the experiment "Dogs & Owners", a substantial experimenter bias can be excluded, since results reported did not confirm the expectations articulated by the author beforehand.

Population sizes: One concern for the internal validity of the gender effects observed could be that they were based on a significantly lower number of female participants compared to the number of male participants that took part in the studies. Unfortunately, the interest among female staff and students to participate in the experiments was low, which made it very hard to find more female volunteers for the limited time frame when the experimental conditions were set up. However, the gender differences observed for the smaller number of female participants were salient, consistent, and substantial, which suggests that their consideration was well-founded.

Furthermore, only a limited number of participants were considered in the transcription analysis conducted in experiment "Celebrity Quote", which lowers the statistical power of the results. However, transcribing dialogues is extremely time consuming – it took six full weeks to transcribe the dialogues of the selected sixteen participants – and was therefore not feasible for the full number of participants.

Chapter 9

Conclusion

Video-mediated communication (VMC) is the current prevalent mode of telecommunication for applications such as remote collaboration, teleconferencing, and distance learning. It is generally assumed that transmitting real-time video of participants in addition to their audio is beneficial and desirable, enabling remote conferencing to feel almost the same as face-to-face collaboration. However, several studies have challenged these assumptions, showing instead that VMC is more similar to telephone conversations than unmediated face-to-face communication.

To improve VMC systems, research has focused on identifying existing shortcomings of standard VMC systems compared to face-to-face communication, and overcoming these limitations through novel interface design. In this context, the interface approach of Video Collaborative Virtual Environments aims to support several spatial aspects of face-to-face meetings that are otherwise absent in standard video collaboration. Spatial conferencing environments are therefore expected to create a stronger sense of co-location, improved gaze-awareness, and a facilitated establishment of a shared collaborative context.

To date, research efforts following this approach have primarily focused on the demonstration of working prototypes. However, maturation of these systems presumes a deeper understanding of the subtle but critical issues that emerge during mediated collaborative processes. The goal of this thesis was to enhance our understanding of these issues by investigating if video-CVEs can effectively support otherwise missing face-to-face aspects of collaboration.

Following the tradition of cross-media studies, the quality of collaboration afforded by video-CVEs was assessed in a relative rather than absolute way, by directly contrasting aspects of the users' experience and collaborative behaviour across multiple conditions, including video-CVEs, standard video-conferencing conditions, and unmediated face-to-face base-line controls. The results obtained were able to indicate benefits and detriments of video-CVEs compared to standard video-conferencing interfaces, and also revealed potential issues and gaps compared to the face-to-face situation.

User experience dimensions included social presence, physical presence, awareness, ease of use, and satisfaction. The collaborative behaviour was assessed and compared using communication patterns, Linguistic features, and view coordination processes. Subjective and objective measures of these dimensions were implemented and applied in four controlled user studies, investigating the impact of two pivotal interface characteristics of video-CVEs: *spatiality*, namely their level of support for fundamental physical spatial properties; and *transportation*, namely the level to which they transport peoples' presence to a remote, artificial space.

9.1 Summary of experiments

9.1.1 Experiment "Desert Survival Game"

The experiment "Desert Survival Game", reported in Chapter 4, was the first experiment comparing spatial with standard video-conferencing. The focus was therefore set on trialling several subjective cross-media metrics for their applicability in the given research focus, while at the same time exploring differences in the communication and interaction patterns observed. Teams of three participants worked on a standard conversational task, the "Desert Survival Game" (DSG), under three experimental conditions: being co-located in a face-to-face situation; being remote, connected through a standard video-conferencing interface; and being remote, connected through a spatial desktop video-CVE interface.

Participants characterised the video-CVE to be the warmer, more personal, more sensitive, and more sociable telecommunication medium, which suggested that higher levels of spatiality and transportation of an interface leads to higher social presence and thus pushes the user experience closer to the face-to-face gold standard. Other subjective measures proved to be too insensitive and were disqualified for further studies.

9.1.2 Experiment "Dream House"

The experiment "Dream House", reported in Chapter 5, sought to confirm the result found in the previous experiment for a different collaborative scenario, and to investigate the applicability of further subjective scales. Teams of two participants worked on a design review task in a standard video-conferencing and in a video-CVE condition. Rating scales assessed physical presence, social presence, as well as preference scores for both interfaces.

The findings showed that both social presence and physical presence were higher in the video-CVE, which confirmed and replicated the result of experiment "Desert Survival Game", and also showed that people indeed perceived their presence to be "transported" to the virtual space provided by the video-CVE. A closer inspection of the relationship between physical and social presence found only a weak correlation. However, the relationship of the team partners was found to have affected the outcome: a stronger correlation between the two constructs was found if team partners were friends. The majority of participants interviewed preferred the video-CVE over the standard video-conferencing interface for collaboration in the given scenarios.

9.1.3 Experiment "Dogs & Owners"

The experiment "Dogs & Owners", reported in Chapter 6, sought to isolate the spatiality aspect of virtual spatial interaction. A range of subjective and objective measures were applied to compare collaboration of two same-gender friends discussing a set of photos across four conditions: two spatial conditions which closely simulated spatial collaboration of two people sitting around a shared table, a standard video-conferencing condition, and a face-to-face condition.

The findings showed that the spatial conditions supported a higher degree of gaze awareness, social presence, and copresence. However, drawbacks regarding usability and task efficiency surfaced, which ultimately led participants to marginally prefer the standard video-conferencing condition over both spatial conditions.

9.1.4 Experiment "Celebrity Quotes"

The experiment "Celebrity Quotes", presented in Chapter 7, focused on the level of "transportation" of a video-CVE. In particular, it sought to determine whether a higher level of transportation of a video-CVE affords a user experience and behaviour that is more similar to face-to-face collaboration. The level of transportation of two video-CVE conditions was controlled by using different types of displays to present the video-CVE to the participants: a standard desktop monitor for the low immersion condition and a large stereo-projection for the high immersion condition. Also included in the experiment was a standard video-conferencing condition as well as a face-to-face control condition. Teams consisting of same-gender friends worked on a collaborative task in each of the conditions. Subjective measures assessed participants' ratings for social presence, physical presence, ease of use, awareness, and preference across the four conditions. Objective measures, based on recorded videos and conversation transcripts, furthermore assessed differences in communication patterns and awareness mechanisms across the four conditions.

The video-CVE with the higher level of transportation induced a stronger sense of physical presence. Other findings were subject to a substantial differences between male and female teams. Male participants rated the more immersive video-CVE as supporting the highest level of social presence and awareness, followed by the less immersive video-CVE, and the standard video-conferencing tool. In contrast to this, female participants rated these systems in exact the opposite order. Participants' assessment of the immersive video-CVE was often subject to confounding factors like motion sickness or technology aversion.

Further gender differences were also exposed by the objective measures in the face-to-face condition. Male participants adopted a "side-by-side" collaborative style, with both participants directed towards a common artefact in the room. This allowed them to maintain a shared focus of attention and enabled them to reduce verbal effort. Female participants, in contrast, favoured "face-to-face" arrangements, where they compromised a shared view of an artefact for being able to look straight at their partner's face.

Based on these observations, it was concluded that the types of collaboration afforded by the different systems to a greater or lesser extent support the different collaborative styles preferred by male or female friends.

9.2 Key questions addressed in this work

The goal of this research was to enhance our understanding of human factors involved in spatial virtual interactions afforded by video-CVEs, and thus inform the iterative improvement of the design of such systems in the future. Several key questions were addressed in this work. The overarching question concerned the main research hypothesis:

> Do geographically dispersed people who collaborate in video-CVEs feel and behave more similarly to being face-to-face than if they collaborate through standard video-conferencing systems?

The next four questions are derived from the first question, but are more focused on the beneficial value of properties of video-CVEs that set them apart from conventional video-conferencing systems. The interface properties that are distinct in video-CVEs were identified in Section 3.2.2 as spatiality and transportation, with video-CVEs having a higher level of spatiality and transportation than conventional video-conferencing interfaces.

Social presence was of interest for assessing the value of spatiality and transportation, since it assesses a subjective attitudinal dimension which is directly related to additional non-verbal channels available in video-CVEs. Communication efficiency was of interest for assessing the value of spatiality and transportation, since it is a key indicator if a telecommunication medium provides useful visual information that allows users to reduce their verbal effort. Both social presence and communication efficiency are highest in face-to-face communication. Differences that emerged between interfaces could therefore indicate if an increase in spatiality and transportation could approximate mediated collaboration to face-to-face collaboration. The other questions of interest were the following:

> Does more spatiality lead to more communication efficiency?

Does more spatiality lead to higher social presence?

Does more transportation lead to more communication efficiency?

Does more transportation lead to higher social presence?

Besides comparing and contrasting collaboration supported by video-CVEs with standard video collaboration, several exploratory questions concerned a better understanding of how people use video-CVEs.

How does an interface influence awareness?

Which usability bottlenecks surface in video-CVEs?

Which usage patterns for video-CVEs emerge?

Which interface type do participants prefer, and why?

What is the relationship between social presence and preference?

Are there differences between strangers and friends and between males and females?

Finally, some questions posed in this research related to the methodological issues.

How can the quality of remote collaboration be assessed?

What experiment scenarios allow useful insights?

Which experiment tasks allow useful insights?

9.3 Main contributions

In the globalised world we live in today, there is a constantly growing demand for the ever-richer telecommunication media that allows dispersed people to meet, talk, and work together in a way that is similar to face-to-face interaction. In this regard, the opportunities offered by avatar-mediated collaboration receive increasing attention.

This thesis contrasted remote collaboration afforded by spatial virtual interaction relative to standard video collaboration as well as face-to-face collaboration and made both methodological and substantive contributions to better understand the human factors involved. The substantive contributions consist of empirical findings of a series of cross-media comparisons involving video-CVEs and standard video-conferencing systems. The methodological contributions concern the research approach taken and the measures explored.

9.3.1 Methodological contributions

Different existing subjective cross-media measures were tested for their applicability in the context of comparing users' interaction experience while collaborating in spatial and non-spatial video-conferencing systems. Researchers who run similar studies in the future will benefit from these practical experiences when selecting from these measurement instruments. Furthermore, rating scales assessing a sense of spatial copresence, the ease of use of a system, and awareness were designed, applied, and validated within the context of this research which may be usefully re-applied in similar settings. 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 measures that were applied in the experiments, the experiment setups themselves contained elements that departed from standard procedure. The experiment "Dream House" introduced a design review task for use in the study of remote collaboration. The experiment "Dogs & Owners" applied a photo-based collaborative task which was especially designed to match the focus and design of the experiment: it could be split into four independent rounds, was not prone to learning effects, and involved a high level of uncertainty, which required rich communication to resolve. The task and scenario used in experiment "Celebrity Quotes" was designed to allow the assessment of verbal effort based on different types of referencing mechanisms that could be extracted from the communication transcripts.

9.3.2 Substantive contributions

The experiments revealed aspects of remote interactions that set video-CVEs apart from standard video-conferencing. Some findings demonstrated where video-CVEs were capable of pushing the telecommunication experience closer to being face-to-face. The assessment of social presence of all media conditions applied in all experiments showed that video-CVEs yield a higher level of social presence than standard video-conferencing systems, making them the warmer, more personal, more sensitive, and more sociable and thus more face-to-face-like telecommunication medium. This finding, however, had to be limited to the group of male participants, since the social presence results were subject to a detected gender effect. A smaller group of female participants saw social presence of standard video-conferencing to be superior. Based on other observations, this gender effect was concluded to be the result of differences in gender-specific communication styles. Video-CVEs also afforded a higher sense of copresence, that is, the feeling of actually being in the same room as a remote partner, which suggests that these systems are capable of bringing about a social context that is more similar to face-to-face.

However, other findings identified areas where avatar-mediated collaboration polled worse than standard video-collaboration and deviated further from face-to-face collaboration. Participants found video-CVEs harder to use than standard video-conferencing systems, mainly because of the introduced need for navigation. The more complex handling of spatial systems introduced a higher level of confusion and uncertainty, which participants resolved verbally. The additional verbal overhead outweighed the benefits gained from the provision of additional non-verbal grounding mechanisms. Consequently, communication efficiency decreased.

These results revealed a tension between the sense of *spatiality* and the *ease of interaction* of tele-collaboration. While video-CVEs created a sense of a social context that was more similar to face-to-face, the standard video-conferencing interfaces supported easier interactions that conformed more with the low effort of face-to-face interactions.

Furthermore, two experiments investigated the relationship between social presence and physical presence. Physical presence enfolds the sense of "being there" in a remote, artificial space, and is therefore closely related to the level of transportation of an interface. The hypothesis was that people would consider collaborative systems that induced a stronger sense of physical presence to be the warmer, more personal, more sociable, and more sensitive telecommunication medium.

The results showed only weak support for this hypothesis. Participants' social presence ratings did not closely reflect their physical presence scores, but rather aligned with their comfort levels during the immersive experience. The comfort levels were in turn affected by confounding factors such as motion sickness, technophobic tendencies, and by the relationship of the participants collaborating.

9.3.3 Minor contributions

Minor theoretical contributions included a simple model for interpersonal collaboration and an in-depth review of cross-media studies classified by different types of measures applied. Minor technical contributions included the implementation of several experiment conditions including a spatial consistent arrangement of two remote participants collaborating around a shared interactive table.

9.4 Conclusion

This research has demonstrated opportunities as well as challenges of supporting remote collaboration with video-CVEs compared to normal video-collaboration and has thus contributed to a better understanding of the human factors involved in spatial virtual interaction. These findings will inform the next step towards richer video-CVEs – richer in value for people using them.

References

Acker, S. and Levitt, S.: 1987, Designing videoconferencing facilities for improved eye contact, *Journal of Broadcasting & Electronic Media* **31**, 181–191.

Argyle, M.: 1992, *The social psychology of everyday life.*, Routledge.

Argyle, M. and Cook, M.: 1976, *Gaze and Mutual Gaze*, Cambridge University Press, UK.

Argyle, M. and Dean, J.: 1965, Eye-contact, distance and affiliation, *Sociometry* **3**, 289–304.

Argyle, M., Lalljee, M. and Cook, M.: 1968, The effects of visibility on interaction in a dyad, *Human Relations* **21**(1), 3–17.

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A. and Blascovich, J.: 2005, 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: Teleoperators & Virtual Environments* **14**(4), 379–393.

Baker, H. H., Bhatti, N., Tanguay, D., Sobel, I., Gelb, D., Goss, M. E., MacCormick, J., Yuasa, K., Culbertson, W. B. and Malzbender, T.: 2003, Computation and performance issues in coliseum: an immersive videoconferencing system, *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, ACM Press, New York, NY, USA, pp. 470–479.

Beaudouin-Lafon, M. and Karsenty, A.: 1992, Transparency and awareness in a real-time groupware system, *UIST '92: Proceedings of the 5th annual*

ACM symposium on User interface software and technology, ACM Press, New York, NY, USA, pp. 171–180.

Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C. and Snowdon, D.: 1995, User embodiment in collaborative virtual environments, *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 242–249.

Benford, S., Bowers, J., Fahlén, L. E. and Mariani, J.: 1994, Supporting cooperative work in virtual environments, *The Computer Journal* **37**(4), 653–668.

Benford, S., Brown, C., Reynard, G. and Greenhalgh, C.: 1996, Shared spaces: transportation, artificiality, and spatiality, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 77–86.

Benford, S. and Fahlén, L.: 1993, A spatial model of interaction in large virtual environments, *Proceedings of the Third European Conference on Computer-Supported Cooperative Work (ESCW '93).*

Benford, S. and Greenhalgh, C.: 1997, Introducing third party objects into the spatial model of interaction, *ECSCW'97: Proceedings of the fifth conference on European Conference on Computer-Supported Cooperative Work*, Kluwer Academic Publishers, Norwell, MA, USA, pp. 189–204.

Benford, S., Greenhalgh, C., Rodden, T. and Pycock, J.: 2001, Collaborative virtual environments, *Commun. ACM* **44**(7), 79–85.

Biocca, F., Harms, C. and Burgoon, J. K.: 2003, Toward a more robust theory and measure of social presence: review and suggested criteria, *Presence: Teleoper. Virtual Environ.* **12**(5), 456–480.

Biocca, F., Harms, C. and Gregg, J.: 2001, The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity, *Presence 2001, 4th international workshop*, Philadelphia, PA.

Bly, S. A.: 1988, A use of drawing surfaces in different collaborative settings, *CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 250–256.

Bly, S. A., Harrison, S. R. and Irwin, S.: 1993, Media spaces: bringing people together in a video, audio, and computing environment, *Commun. ACM* **36**(1), 28–46.

Borning, A. and Travers, M.: 1991, Two approaches to casual interaction over computer and video networks, *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 13–19.

Bortz, J. and Döring, N.: 2003, *Forschungsmethoden und Evaluation*, Springer.

Bos, N., Olson, J., Gergle, D., Olson, G. and Wright, Z.: 2002, Effects of four computer-mediated communications channels on trust development, *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 135–140.

Bowers, J., Pycock, J. and O'Brien, J.: 1996, Talk and embodiment in collaborative virtual environments, *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 58–65.

Brown, B. and Bell, M.: 2004, Cscw at play:''there' as a collaborative virtual environment, *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 350–359.

Buxton, B. and Moran, T.: 1990, Europarc's integrated interactive intermedia facility (iiif): Early experiences, *in* S. Giggs and Verrijn-Stuart (eds), *Proceedings of the IFIP WG 8.4 Conference on Multi-user Interfaces and Applications*, Elsevier Science Publishers B.V. (North-Holland), pp. 11–34.

Carey, J.: 1989, *Communication as Culture*, New York: Routledge.

Carlsson, C. and Hagsand, O.: 1993, Dive - a platform for multi-user virtual environments, *Computer and Graphics* **17**(6), 663–669.

Chapanis, A.: 1975, Interactive human communication, *Scientific American* **232**(3), 36–42.

Chen, W.-C., Towles, H., Nyland, L., Welch, G. and Fuchs, H.: 2000, Toward a compelling sensation of telepresence: Demonstrating a portal to a distant (static) office, *VISUALIZATION '00: Proceedings of the 11th IEEE Visualization 2000 Conference (VIS 2000)*, IEEE Computer Society, Washington, DC, USA.

Clark, H. H.: 1985, Language use and language users, *in* G. Lindzey and E. Aronson (eds), *Handbook of social psychology*, 3rd ed. edn, New York: Harper and Row, pp. 179–231.

Clark, H. H. and Brennan, S. E.: 1991, Grounding in communication, *in* L. B. Resnick, J. M. Levine and S. D. Teasley (eds), *Perspectives on socially shared cognition*, Washington: APA Books, pp. 127–149.

Clark, H. H. and Wilkes-Gibbs, D.: 1986, Referring as a collaborative process, *Cognition* **22**, 1–39.

Cockburn, A. and Greenberg, S.: 1993, Making contact: getting the group communicating with groupware, *COCS '93: Proceedings of the conference on Organizational computing systems*, ACM, New York, NY, USA, pp. 31–41.

Coelho, C., Tichon, J., Hine, T. J., Wallis, G. and Riva, G.: 2006, Media presence and inner presence: The sense of presence in virtual reality technologies, *in* G. Riva, B. W. M.T. Anguera and F. Mantovani (eds), *From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience. Festschrift in honor of Luigi Anolli*, IOS Press, Amsterdam, pp. 25–45.

Cool, C., Fish, R. S., Kraut, R. E. and Lowery, C. M.: 1992, Iterative design of video communication systems, *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 25–32.

Crabtree, A., Rodden, T. and Mariani, J.: 2004, Collaborating around collections: informing the continued development of photoware, *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 396–405.

Daft, R. L., Lengel, R. and Trevino, L. K.: 1987, Message equivocality, media selection, and manager performance: Implications for information systems, *MIS Quarterly* **11**(3), 355–366.

Daft, R. and Lengel, R.: 1984, Information richness: A new approach to managerial behavior and organization design, *Research in Organizational Behavior* **6**, 191–233.

Daft, R. and Lengel, R.: 1986, Organizational information requirements, media richness, and structural design, *Management Science* **32**, 554–571.

Daly-Jones, O., Monk, A. and Watts, L.: 1998, Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus, *Int. J. Hum.-Comput. Stud.* **49**(1), 21–58.

de Greef, P. and IJsselsteijn, W. A.: 2001, Social presence in a home tele-application, *Cyberpsychology & Behavior* **4**(2), 307–315.

Denis, A. R. and Kinney, S. T.: 1998, Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality, *Information Systems Research* **9**(3), 256–274.

DeVito, J. A.: 1998, *The interpersonal communication book*, 8th edn, New York: Longman.

Dewan, P.: 1999, Architectures for collaborative applications, *in* M. Beaudouin-Lafon (ed.), *Computer Supported Co-operative Work*, John Wiley and sons, pp. 169–193.

Doherty-Sneddon, G., O'Malley, C., Garrod, S., Anderson, A., Langton, S. and Bruce, V.: 1997, Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance, *Journal of Experimental Psychology: Applied* **3**(2), 105–125.

Douglas, A.: 1957, The peaceful settlement of industrial and intergroup disputes, *Journal of Conflict Resolution* **1**(1), 69–81.

Dourish, P.: 2006, Re-space-ing place: "place" and "space" ten years on, *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 299–308.

Dourish, P. and Bellotti, V.: 1992, Awareness and coordination in shared workspaces, *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 107–114.

Dourish, P. and Bly, S.: 1992, Portholes: supporting awareness in a distributed work group, *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 541–547.

Egido, C.: 1988, Video conferencing as a technology to support group work: a review of its failures, *CSCW '88: Proceedings of the 1988 ACM confer-*

ence on Computer-supported cooperative work, ACM Press, New York, NY, USA, pp. 13–24.

El-Shinnawy, M. and Markus, M. L.: 1998, Acceptance of communication media in organizations: richness or features, *IEEE Transactions on Professional Communication* **41**(4), 242–253.

Engelbart, D. C. and English, W. K.: 1968, A research center for augmenting human intellect, *AFIPS Conference Proceedings*, Vol. 33.

Fish, R. S., Kraut, R. E. and Chalfonte, B. L.: 1990, The videowindow system in informal communication, *CSCW '90: Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 1–11.

Fish, R. S., Kraut, R. E., Root, R. W. and Rice, R. E.: 1992, Evaluating video as a technology for informal communication, *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 37–48.

Fraser, M., Benford, S., Hindmarsh, J. and Heath, C.: 1999, Supporting awareness and interaction through collaborative virtual interfaces, *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, ACM, New York, NY, USA, pp. 27–36.

Frécon, E. and Smith, G.: 1999, Semantic behaviours in collaborative virtual environments, *Proceedings of 5th Eurographics on Virtual Environments*, pp. 95–104.

Frécon, E., Smith, G., Steed, A., Stenius, M. and Stahl, O.: 2001, An overview of the coven platform, *Presence: Teleoper. Virtual Environ.* **10**(1), 109–127.

Frécon, E. and Stenius, M.: 1998, Dive: A scalable network architecture for distributed virtual environments, *Distributed Systems Engineering Journal* **5**(3), 91–100.

Friedman, D., Leeb, R., Guger, C., Steed, A., Pfurtscheller, G. and Slater, M.: 2007, Navigating virtual reality by thought: what is it like?, *Presence: Teleoper. Virtual Environ.* **16**(1), 100–110.

Fussell, S. R., Kraut, R. E. and Siegel, J.: 2000, Coordination of communication: effects of shared visual context on collaborative work, *CSCW '00: Proceedings of the 2000 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 21–30.

Fussell, S. R., Setlock, L. D. and Kraut, R. E.: 2003, Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks, *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 513–520.

Gale, C. and Monk, A. F.: 2000, Where am i looking? the accuracy of video-mediated gaze awareness, *Perception & Psychology* **62**(3), 586–595.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A. and Sasse, M. A.: 2003, The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment, *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 529–536.

Gaver, W., Moran, T., MacLean, A., Lövstrand, L., Dourish, P., Carter, K. and Buxton, W.: 1992, Realizing a video environment: Europarc's rave system, *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 27–35.

Gaver, W. W.: 1992, The affordances of media spaces for collaboration, *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 17–24.

Gaver, W. W., Sellen, A., Heath, C. and Luff, P.: 1993, One is not enough: multiple views in a media space, *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 335–341.

Gibbs, S. J., Arapis, C. and Breiteneder, C. J.: 1999, Teleport towards immersive copresence, *Multimedia Syst.* **7**(3), 214–221.

Gibson, J. J.: 1979, *The ecological approach to visual perception*, Houghton Mifflin, New York.

Goodwin, C. J.: 1995, *Research in Psychology*, John Wiley & Sons, Inc.

Gouran, D., Wiethoff, W. E. and Doelger, J. A.: 1994, *Mastering communication*, Allyn and Bacon (Boston).

Greenhalgh, C. and Benford, S.: 1995, Massive: a collaborative virtual environment for teleconferencing, *ACM Trans. Comput.-Hum. Interact.* **2**(3), 239–261.

Griffin, E. A.: 2005, *A first look at communication theory*, 5th edn, McGraw-Hill.

Gueddana, S. and Roussel, N.: 2006, Pêle mêle, a video communication system supporting a variable degree of engagement, *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 423–426.

Gutwin, C.: 1997, *Workspace Awareness in Real-Time Distributed Groupware*, PhD thesis, University of Calgary.

Gutwin, C. and Greenberg, S.: 1996, Workspace awareness for groupware, *CHI '96: Conference companion on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 208–209.

Gutwin, C. and Greenberg, S.: 1998, Design for individuals, design for groups: tradeoffs between power and workspace awareness, *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 207–216.

Gutwin, C. and Greenberg, S.: 1999, The effects of workspace awareness support on the usability of real-time distributed groupware, *ACM Trans. Comput.-Hum. Interact.* **6**(3), 243–281.

Hall, E. T.: 1969, *The hidden Dimension*, Anchor books, Doubleday & Company, Inc.

Hall, J. A.: 1990, *Nonverbal Sex Differences*, The Johns Hopkins University Press;.

Han, J. and Smith, B.: 1996, Cu-seeme vr immersive desktop teleconferencing, *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia*, ACM Press, New York, NY, USA, pp. 199–207.

Hansen, K.: 2002, The design of public virtual spaces in 3d virtual worlds on the internet, *in* L. Qvortrup (ed.), *Virtual Space: Spatiality in Virtual Inhabilted 3D Worlds*, Springer London, pp. 145–170.

Harrison, S. and Dourish, P.: 1996, Re-place-ing space: the roles of place and space in collaborative systems, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 67–76.

Harwood, J.: 2000, Communication media use in the grandparent-grandchild relationship, *Journal of Communication* **50**, 56–78.

Hauber, J., Billinghurst, M. and Regenbrecht, H.: 2004, Tangible teleconferencing, *Proceedings of 6th Asia Pacific Conference, APCHI 2004*, Springer Berlin / Heidelberg, pp. 143–152.

Heath, C. and Luff, P.: 1991, Disembodied conduct: communication through video in a multi-media office environment, *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 99–103.

Heeter, C.: 1992, Being there: The subjective experience of presence., *Presence: Teleoper. Virtual Environ.* **1**(2), 262–271.

Held, R. M. and Durlach, N. I.: 1992, Telepresence, *Presence: Teleoper. Virtual Environ.* **1**(1), 109–112.

Hills, A. L.: 2005, Social presence and communication quality in videoconferencing, Dissertation.

Hornecker, E.: 2004, *Tangible User Interfaces als kooperationsunterstuetzendes Medium*, PhD thesis, University of Bremen.

Hudson-Smith, A.: 2002, 30 days in active worlds: Community, design and terrorism in a virtual world, *in* R. Schroeder (ed.), *The Social Life of Avatars*, Springer, pp. 77–89.

IJsselsteijn, W.: 2003, Presence in the past: what can we learn from media history?, *in* W. I. G. Riva, F. Davide (ed.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*, Ios Press, pp. 17–40.

IJsselsteijn, W. A., Freeman, J. and de Ridder, H.: 2001, Presence: Where are we?, *Cyberpsychology & Behavior* **4**(2), 179–182.

IJsselsteijn, W. and Riva, G.: 2003, Being there: The experience of presence in mediated environments., *in* W. I. G. Riva, F. Davide (ed.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*, Ios Press, pp. 4–14.

Insko, B.: 2001, *Passive Haptics Significantly Enhance Virtual Environments*, PhD thesis, University of North Carolina at Chapel Hill.

Insko, B. E.: 2003, Measuring presence: Subjective, behavioral, and physiological methods., *in* W. I. G. Riva, F. Davide (ed.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*, Ios Press, pp. 110–118.

Ishii, H. and Kobayashi, M.: 1992, Clearboard: a seamless medium for shared drawing and conversation with eye contact, *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 525–532.

Ishii, H., Kobayashi, M. and Arita, K.: 1994, Iterative design of seamless collaboration media, *Commun. ACM* **37**(8), 83–97.

Ishii, H., Kobayashi, M. and Grudin, J.: 1993, Integration of interpersonal space and shared workspace: Clearboard design and experiments, *ACM Trans. Inf. Syst.* **11**(4), 349–375.

Ishii, H. and Ullmer, B.: 1997, Tangible bits: towards seamless interfaces between people, bits and atoms, *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, New York, NY, USA, pp. 234–241.

Johansen, R.: 1988, *GroupWare: Computer Support for Business Teams*, The Free Press.

Jouppi, N. P.: 2002, First steps towards mutually-immersive mobile telepresence, *CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 354–363.

Kauff, P. and Schreer, O.: 2002, An immersive 3d video-conferencing system using shared virtual team user environments, *CVE '02: Proceedings of the 4th international conference on Collaborative virtual environments*, ACM Press, New York, NY, USA, pp. 105–112.

Kendon, A.: 1967, Some functions of gaze-direction in social interaction, *Acta Psychologica* **26**, 22–63.

Kramer, A. D. I., Oh, L. M. and Fussell, S. R.: 2006, Using linguistic features to measure presence in computer-mediated communication, *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM Press, New York, NY, USA, pp. 913–916.

Kraut, R. E., Miller, M. D. and Siegel, J.: 1996, Collaboration in performance of physical tasks: effects on outcomes and communication, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 57–66.

Kraut, R. E., Rice, R. E., Cool, C. and Fish, R. S.: 1994, Life and death of new technology: task, utility and social influences on the use of a communication medium, *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 13–21.

Kruger, R., Carpendale, S., Scott, S. D. and Greenberg, S.: 2004, Roles of orientation in tabletop collaboration: Comprehension, coordination and communication, *Comput. Supported Coop. Work* **13**(5-6), 501–537.

Lafferty, J. C., Eady, P. M. and Elmers, J.: 1974, The desert survival problem, *Experimental Learning Methods*, Plymouth, Michigan.

Lasswell, H.: 1948, The structure and function of communication in society, *in* L. Bryson (ed.), *The Communication of Ideas*, New York: Institute for Religious and Social Studies.

Lauwers, J. C., Joseph, T. A., Lantz, K. A. and Romanow, A. L.: 1990, Replicated architectures for shared window systems: a critique, *Proceedings of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems*, ACM Press, New York, NY, USA, pp. 249–260.

Lee, A. S.: 1994, Electronic mail as a medium for rich communication: An empirical investigation using hermeneutic interpretation, *MIS Quarterly* **18**(2), 143–157.

Leeb, R., Keinrath, C., Friedman, D., Guger, C., Scherer, R., Neuper, C., Garau, M., Antley, A., Steed, A., Slater, M. and Pfurtscheller, G.: 2006, Walking by thinking: the brainwaves are crucial, not the muscles!, *Presence: Teleoper. Virtual Environ.* **15**(5), 500–514.

Lessiter, J., Freeman, J., Keogh, E. and Davidoff, J.: 2001, A cross-media presence questionnaire: The itc-sense of presence inventory, *Presence: Teleoper. Virtual Environ.* **10**(3), 282–297.

Lombard, M. and Ditton, T.: 1997, At the heart of it all: The concept of presence., *Journal of Computer Mediated Communication* **3**(2).

Looser, J.: 2007, *Analysing Videos with VideoAnalysisApp.*

Luce, R. and Raiffa, H.: 1957, *Games and Decisions*, Wiley: New York.

Mackay, W. E.: 1999, Media spaces: Environments for informal multimedia interaction, *in* M. Beaudouin-Lafon (ed.), *Computer Supported Cooperative Work*, John Wiley and sons, pp. 55–82.

Manly, B. F.: 1986, *Multivariate statistical methods: A primer*, London : Chapman and Hall.

Mantovani, F., Agliati, A., Mortillaro, M., Vescovo, A. and Zurloni, V.: 2006, Communication - presence roundtrip: Travelling along theoretical, methodological and applicative connections, *in* G. Riva, B. W. M.T. Anguera and F. Mantovani (eds), *From Communication to Presence: Cognition, Emotions and Culture towards the Ultimate Communicative Experience. Festschrift in honor of Luigi Anolli*, IOS Press, Amsterdam, pp. 3–22.

Mason, R.: 1994, *Using communciations media in open and flexible learning.*, London: Kogan Page.

McGrath, A. and Prinz, W.: 2001, All that is solid melts into software, *in* E. F. Churchill, D. N. Snowdon and A. J. Munro (eds), *Collaborative Virtual Environments*, Springer, pp. 99–114.

Monk, A. F. and Gale, C.: 2002, A look is worth a thousand words: Full gaze awareness in video-mediated conversation, *Discourse Processes* **33**(3), 257–278.

Monk, A., McCarthy, J., Watts, L. and Daly-Jones, O.: 1996, Evaluation for cscw, *in* P. Thomas (ed.), *Measures of Process*, Springer, pp. 125–139.

Morikawa, O. and Maesako, T.: 1998, Hypermirror: toward pleasant-to-use video mediated communication system, *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 149–158.

Morley, I. and Stephenson, G.: 1969, Interpersonal and interparty exchange. a laboratory simulation of an industrial negotiation at the plant level, *British Journal of Psychology* **60**, 543–545.

Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: 1996, Freewalk: supporting casual meetings in a network, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 308–314.

Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: 1998, Freewalk: A three-dimensional meeting-place for communities, *in* T. Ishida (ed.), *Community Computing: Collaboration over Global Information Networks*, John Wiley and Sons, pp. 55–89.

Nardi, B. A., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S. and Sclabassi, R.: 1993, Turning away from talking heads: the use of video-as-data in neurosurgery, *CHI '93: Proceedings of the SIGCHI conference* *on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 327–334.

Nardi, B. and Harris, J.: 2006, Strangers and friends: collaborative play in world of warcraft, *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 149–158.

Nash, E. B., Edwards, G. W., Thompson, J. A. and Barfield, W.: 2000, A review of presence and performance in virtual environments, *International Journal of Human-Computer Interaction* **12**(1), 1–41.

Nguyen, D. and Canny, J.: 2005, Multiview: spatially faithful group video conferencing, *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 799–808.

Normand, V., Babski, C., Benford, S., Bullock, A., Carion, S., Chrysanthou, Y., Farcet, N., Frecon, E., Harvey, J., Kuijpers, N., Magnenat-Thalmann, N., Raupp-Musse, S., Rodden, T., Slater, M., Smith, G., Steed, A., Thalmann, D., Tromp, J., Usoh, M., Liempd, G. V. and Kladias, N.: 1999, The coven project: Exploring applicative, technical, and usage dimensions of collaborative virtual environments, *Presence: Teleoper. Virtual Environ.* **8**(2), 218–236.

Nowak, K.: 2001, Defining and differentiating copresence, social presence and presence as transportation, *PRESENCE 2001 - 4th annual international workshop*.

Nowak, K. L. and Biocca, F.: 2003, The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments, *Presence: Teleoper. Virtual Environ.* **12**(5), 481–494.

O'Conaill, B., Whittaker, S. and Wilbur, S.: 1993, Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication, *Human-Computer Interaction* **8**(4), 389–428.

Okada, K.-I., Maeda, F., Ichikawaa, Y. and Matsushita, Y.: 1994, Multi-party videoconferencing at virtual social distance: Majic design, *CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 385–393.

Olson, G. M. and Olson, J. S.: 2000, Distance matters, *Human Computer Interaction* **15**, 139–179.

Olson, J. S., Card, S. K., Landauer, T. K., Olson, G. M., Malone, T. and Leggett, J.: 1993, Computer-supported co-operative work: research issues for the 90s., *Behaviour & Information Technology* **12**(2), 115–129.

Olson, J. S., Olson, G. M. and Meader, D. K.: 1995, What mix of video and audio is useful for small groups doing remote real-time design work?, *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 362–368.

Olson, M. H. and Bly, S. A.: 1991, The portland experience: a report on a distributed research group, pp. 81–98.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G. and Bruce, V.: 1996, Comparison of face-to-face and video-mediated interaction., *Interacting with Computers* **8**(2), 177–192.

Osgood, C. E., Suci, G. J. and Tannenbaum, P. H.: 1957, *The measurement of meaning*, Urbana.

Parise, S., Kiesler, S., Sproull, L. and Waters, K.: 1996, My partner is a real dog: cooperation with social agents, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 399–408.

Pennebaker, J. W., Francis, M. E. and Booth, R. J.: 2003, Linguistic inquiry and word count: Liwc2001, Manual.

Prakash, A.: 1999, *Computer Supported Co-operative Work*, John Wiley and Sons, chapter Group Editors, pp. 104–133.

Prinz, W.: 1999, Nessie: an awareness environment for cooperative settings, *ECSCW'99: Proceedings of the sixth conference on European Conference on Computer Supported Cooperative Work*, Kluwer Academic Publishers, Norwell, MA, USA, pp. 391–410.

Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L. and Fuchs, H.: 1998, The office of the future: a unified approach to image-based modeling and spatially immersive displays, *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, pp. 179–188.

Reddy, M. J.: 1979, The conduit metaphor: A case of frame conflict in our language about langauge., *in* A. Ortony (ed.), *Metaphor and Thought*, Cambridge: Cambridge University Press, pp. 284–297.

Regenbrecht, H., Haller, M., Hauber, J. and Billinghurst, M.: 2006, Carpeno: interfacing remote collaborative virtual environments with table-top interaction, *Virtual Real.* **10**(2), 95–107.

Regenbrecht, H., Lum, T., Kohler, P., Ott, C., Wagner, M., Wilke, W. and Mueller, E.: 2004, Using augmented virtuality for remote collaboration, *Presence: Teleoper. Virtual Environ.* **13**(3), 338–354.

Reitman, W. R.: 1965, *Cognition and Thought: An Information-Processing Approach*, Wiley.

Rekimoto, J., Ullmer, B. and Oba, H.: 2001, Datatiles: a modular platform for mixed physical and graphical interactions, *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, New York, NY, USA, pp. 269–276.

Rice, R. E.: 1992, Task analyzability, use of new media, and effectiveness: A multi-site exploration of media richness., *Organization Science* **3**(4), 475–500.

Rice, R. E., D'Ambra, J. and More., E.: 1998, Cross-culture comparsion of organizational media evaluation and choice, *Journal of Communication* **48**(3), 3–26.

Riva, G.: 1999, Virtual reality as communication tool: A sociocognitive analysis, *Presence: Teleoper. Virtual Environ.* **8**(4), 462–468.

Root, R. W.: 1988, Design of a multi-media vehicle for social browsing, *CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, ACM Press, New York, NY, USA, pp. 25–38.

Roseman, M. and Greenberg, S.: 1996, Building real-time groupware with groupkit, a groupware toolkit, *ACM Trans. Comput.-Hum. Interact.* **3**(1), 66–106.

Roszak, T.: 1986, *The cult of information: the folklore of computers and the true art of thinking.*, University of California Press;.

Roy, M. M. and Christenfeld, N. J.: 2004, Do dogs resemble their owners?, *Psychological Sciences* **15**(5), 361–363.

Rutter, D. and Stephenson, G.: 1977, The role of visual communication in synchronising conversation, *European Journal of Social Psychology* **7**(1), 29–37.

Rutter, D., Stephenson, G. and Dewey, M.: 1981, Visual communication and the content and style of conversation, *British Journal of Social Psychology* **20**, 41–52.

Sadowski, W. and Stannay, K.: 2002, Presence in virtual environments, *Handbook of Virtual Environments: Design, Implementation, and Applications*, Lawrence Erlbaum associates, pp. 791–806.

Sallnäs, E.-L.: 2005, Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments, *Presence: Teleoper. Virtual Environ.* **14**(4), 434–449.

Sauer, J.: 2001, Virtual reality in der produktentwicklung, *VDI-Berichte 1614*, VDI Verlag Duesseldorf, Germany.

Schnaedelbach, H., Penn, A., Steadman, P., Benford, S., Koleva, B. and Rodden, T.: 2006, Moving office: Inhabiting a dynamic building, *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 413–422.

Schroeder, R.: 1996, *Possible Worlds: The Social Dynamic of Virtual Reality Technology*, Westview Press, Colorado/Oxford.

Schubert, T., Friedmann, F. and Regenbrecht, H.: 2001, The experience of presence: Factor analytic insights, *Presence: Teleoper. Virtual Environ.* **10**(3), 266–281.

Scott, S. D., Grant, K. D. and Mandryk, R. L.: 2003, System guidelines for co-located, collaborative work on a tabletop display, *ECSCW'03: Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, Kluwer Academic Publishers, Norwell, MA, USA, pp. 159–178.

Sellen, A. J.: 1992, Speech patterns in video-mediated conversations, *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 49–59.

Sellen, A. J.: 1995, Remote conversations: The effects of mediating talk with technology, *Human-Computer Interaction* **10**, 401–444.

Shannon, C. E. and Weaver, W.: 1949, *The Mathematical Theory of Communication*, The University of Illinois Press: Urbana.

Sheridan, T. B.: 1992, Musings on telepresence and virtual presence, *Presence: Teleoper. Virtual Environ.* **1**(1), 120–126.

Short, J. A.: 1974, Effects of medium of communication on experimental negotiation, *Human Relations* **27**(3), 225–234.

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

Slater, M.: 2004, How colorful was your day? why questionnaires cannot assess presence in virtual environments, *Presence: Teleoper. Virtual Environ.* **13**(4), 484–493.

Slater, M., Sadagic, A., Usoh, M. and Schroeder, R.: 2000, Small-group behaviour in a virtual and real environment., *Presence: Teleoper. Virtual Environ.* **9**(1), 37–51.

Slater, M. and Usoh, M.: 1994, Body centered interaction in immersive virtual environments., *in* N. M. Thalmann and D. Thalmann (eds), *Artificial Life and Virtual Reality*, John Wiley and Sons, pp. 125–148.

Smith, G.: 1996, Cooperative virtual environments: lessons from 2d multi user interfaces, *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, ACM Press, New York, NY, USA, pp. 390–398.

Stefik, M., Bobrow, D. G., Foster, G., Lanning, S. and Tatar, D.: 1987, Wysiwis revised: early experiences with multiuser interfaces, *ACM Trans. Inf. Syst.* **5**(2), 147–167.

Stephenson, N.: 1992, *Snow Crash*, Bantam Books (USA).

Tang, J. C.: 1991, Findings from observational studies of collaborative work, *Int. J. Man-Mach. Stud.* **34**(2), 143–160.

Tang, J. C.: 1992, Why do users like video? studies of multimedia-supported collaboration, *Technical report*, Mountain View, CA, USA.

Veinott, E. S., Olson, J., Olson, G. M. and Fu, X.: 1999, Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other, *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 302–309.

Vertegaal, R.: 1999, The gaze groupware system: mediating joint attention in multiparty communication and collaboration, *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 294–301.

Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C.: 2003, Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction, *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, New York, NY, USA, pp. 521–528.

Waldner, M., Hauber, J., Zauner, J., Haller, M. and Billinghurst, M.: 2006, Tangible tiles: design and evaluation of a tangible user interface in a collaborative tabletop setup, *OZCHI '06: Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments*, ACM, New York, NY, USA, pp. 151–158.

Whittaker, S.: 1995, Rethinking video as a technology for interpersonal communications: theory and design implications., *Human-Computer Studies* **42**, 501–529.

Whittaker, S., Geelhoed, E. and Robinson, E.: 1993, Shared workspaces: how do they work and when are they useful?, *Int. J. Man-Mach. Stud.* **39**(5), 813–842.

Wichman, H.: 1970, Effects of isolation and communication on cooperation in a two-person game, *Journal of Personality and Social Psychology* **16**(1), 114–120.

Wiener, N.: 1948, *Cybernetics, or, Control and communication in the animal and the machine*, New York : Wiley.

Williams, E.: 1975, Coalition formation over telecommunication media, *European Journal of Social Psychology* **5**, 503–507.

Williams, E.: 1977, Experimental comparisons of face-to-face and mediated communication: A review, *Psychological Bulletin* **84**(5), 963–976.

Witmer, B. G. and Singer, M. J.: 1998, Measuring presence in virtual environments: A presence questionnaire, *Presence: Teleoper. Virtual Environ.* **7**(3), 225–240.

Wright, P. H.: 1982, Men's friendships, women's friendships and the alleged inferiority of the latter, *Sex Roles* **8**, 1–20.

Appendix A

Questionnaires

A.1 Desert Survival Game

Date ________ / Group _____ / Person ____ / Condition ________

Questionnaire

Indicate your preferred answer by marking an "X" in the appropriate box of the seven-point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply.

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

Statement	☐	☐	☐	☐	☐	☐	☐
1. The other individuals were influenced by my moods.	strongly disagree			undecided			strongly agree
2. I understood what the others meant.	strongly disagree			undecided			strongly agree
3. I was often aware of others in the environment. *"Others" refers to other team members in the communication environment.*	strongly disagree			undecided			strongly agree
4. My behaviour was in direct response to the others' behaviour.	strongly disagree			undecided			strongly agree
5. What I did affected what the others did.	strongly disagree			undecided			strongly agree
6. I was influenced by my partners' moods.	strongly disagree			undecided			strongly agree
7. My thoughts were clear to my partners.	strongly disagree			undecided			strongly agree
8. I sometimes pretended to pay attention to the other individuals.	strongly disagree			undecided			strongly agree

9. I think the other individuals often felt alone.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
10. What the others did affected what I did.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
11. The others' actions were dependent on my actions.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
12. My partners worked with me to complete the task.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
13. I could not act without the others.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
14. The others understood what I meant.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
15. My opinions were clear to the others.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
16. The others' mood did NOT affect my mood/emotional-state.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
17. My actions were dependent on the others' actions.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
18. I was easily distracted when other things were going on around me (in the real room).	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
19. My partners did not help me very much.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
20. The others could not act without me	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree

21. Others were often aware of me in the room. *"Others" refers to other team members in the communication environment (room).*	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
22. The other individuals sometimes pretended to pay attention to me.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
23. The other individuals didn't notice me in the room.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
24. The other individuals tended to ignore me.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
25. When the others were happy, I was happy.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
26. My mood did NOT affect the others' mood/emotional-state.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
27. The other individuals' thoughts were clear to me.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
28. I tended to ignore the other individuals.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
29. I did not help the others very much.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
30. My partners were easily distracted when other things were going on around us (in the real room).	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
31. The opinions of the others were clear.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
32. I paid close attention to the other individuals.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree

33. I worked with the other individuals to complete the task.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
34. The behaviour of the others were in direct response to my behaviour.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
35. When I was happy, the others were happy.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
36. I hardly noticed another individual.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
37. The other individuals paid close attention to me.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
38. I often felt as if I was all alone.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
39. There was a great sense of realism in the environment.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
40. I had a good feel of the people at the other end.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
41. It was like a face-to-face meeting.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree
42. I felt as we were all in the same room.	☐ ☐ ☐ ☐ ☐ ☐ ☐ strongly disagree … undecided … strongly agree

PLEASE RATE ACCORDING TO ATTRIBUTES GIVEN ON THE SCALES:

43. The team's task performance in the activity was ...	☐ poor	☐	☐	☐ average	☐	☐	☐ excellent
44. The task performance was highly affected by the communication medium.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
45. The overall flow of communication between the team was ...	☐ poor	☐	☐	☐ average	☐	☐	☐ excellent
46. I waited for my turn to speak.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
47. The others waited till it was their turn to speak.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
48. Turn taking was complicated.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
49. We were talking over one another.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
50. I felt like I was interrupted.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
51. I interrupted the others.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
52. I was confused.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
53. The others seemed to be confused.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often
54. There was a lot of time when no-one spoke at all.	☐ never	☐	☐	☐ sometimes	☐	☐	☐ often

55. I used body language (such as facial expressions, hand movements) to express myself.	☐ never ☐ ☐ ☐ sometimes ☐ ☐ ☐ often
56. The others used body language to express themselves.	☐ never ☐ ☐ ☐ sometimes ☐ ☐ ☐ often

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

57.	☐ cold	☐	☐	☐	☐	☐	☐ warm
58.	☐ insensitive	☐	☐	☐	☐	☐	☐ sensitive
59.	☐ small	☐	☐	☐	☐	☐	☐ large
60.	☐ ugly	☐	☐	☐	☐	☐	☐ beautiful
61.	☐ impersonal	☐	☐	☐	☐	☐	☐ personal
62.	☐ colourless	☐	☐	☐	☐	☐	☐ colourful
63.	☐ closed	☐	☐	☐	☐	☐	☐ open
64.	☐ passive	☐	☐	☐	☐	☐	☐ active
65.	☐ unsociable	☐	☐	☐	☐	☐	☐ sociable

A.2 Dream House

Date ________ / Group _____ / Person ____ / Condition ________

Questionnaire

Indicate your preferred answer by marking an "X" in the appropriate box of the scale. Please consider the entire scale when making your responses, as the intermediate levels may apply.

Please read the statements below and rate according to the attributes on the *5-point* scale

1.	I had a sense of acting in the virtual space, rather than operating something from the outside.	☐ fully disagree	☐	☐	☐	☐ fully agree
2.	How much did your experience in the virtual environment seem consistent with your real world experience?	☐ not consistent	☐	☐ moderately consistent	☐	☐ very consistent
3.	How real did the virtual world seem to you?	☐ completely real	☐	☐	☐	☐ not real at all
4.	I did not feel present in the virtual space	☐ did not feel present	☐	☐	☐	☐ felt present
5.	I was not aware of my real environment.	☐ fully disagree	☐	☐	☐	☐ fully agree
6.	In the computer generated world I had a sense of "being there"	☐ not at all	☐	☐	☐	☐ very much

7.	Somehow I felt that the virtual world surrounded me	☐ fully disagree	☐	☐	☐	☐ fully agree
8.	I felt present in the virtual space	☐ fully disagree	☐	☐	☐	☐ fully agree
9.	I still paid attention to the real environment.	☐ fully disagree	☐	☐	☐	☐ fully agree
10.	The virtual world seemed more realistic than the real world	☐ fully disagree	☐	☐	☐	☐ fully agree
11.	I felt like I was just perceiving pictures	☐ fully disagree	☐	☐	☐	☐ fully agree
12.	I was completely captivated by the virtual world	☐ fully disagree	☐	☐	☐	☐ fully agree

I rate the type of medium I just used to collaborate with others as:

13.	☐ cold	☐	☐	☐	☐	☐	☐ warm
14.	☐ insensitive	☐	☐	☐	☐	☐	☐ sensitive
15.	☐ small	☐	☐	☐	☐	☐	☐ large
16.	☐ impersonal	☐	☐	☐	☐	☐	☐ personal

17.	closed ☐	☐	☐	☐	☐	☐	☐ open	
18.	passive ☐	☐	☐	☐	☐	☐	☐ active	
19.	unsociable ☐	☐	☐	☐	☐	☐	☐ sociable	
20.	constricted ☐	☐	☐	☐	☐	☐	☐ spacious	
21.	boring ☐	☐	☐	☐	☐	☐	☐ interesting	
22.	dehumanising ☐	☐	☐	☐	☐	☐	☐ humanising	

A.3 Dogs & Owners

Group: ______________ Condition: ______________ Date:____________

Questionnaire

Indicate your preferred answer by marking an "X" in the appropriate box of the seven-point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply.

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

Q1	I think that we have managed to do the task very well.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q2	I was always aware that my partner and I were at different locations.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q3	It was very easy to make myself understood.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q4	There was a lot of time when no-one spoke at all.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q5	I was often confused.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q6	I could always clearly hear my partner's voice.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q7	When I looked at my partner, I could always clearly see his or her face.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q8	I could easily tell where my partner was pointing at.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree

Q9	We were never talking over one another.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q10	I was always aware of my partner's presence.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q11	It was just like being face to face with my partner.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q12	I could not contribute anything to the solution we came up with.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q13	I hardly looked at my partner's face.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q14	I knew exactly when it was my turn to speak.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q15	I could easily tell where my partner was looking.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree
Q16	It felt as if my partner and I were in the same room.	☐ strongly disagree	☐	☐	☐ undecided	☐	☐	☐ strongly agree

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

Q17	☐ cold	☐	☐	☐	☐	☐	☐ warm
Q18	☐ insensitive	☐	☐	☐	☐	☐	☐ sensitive

Q19	☐ small	☐	☐	☐	☐	☐	☐ large
Q20	☐ spontaneous	☐	☐	☐	☐	☐	☐ formal
Q21	☐ impersonal	☐	☐	☐	☐	☐	☐ personal
Q22	☐ passive	☐	☐	☐	☐	☐	☐ active
Q23	☐ unsociable	☐	☐	☐	☐	☐	☐ sociable
Q24	☐ open	☐	☐	☐	☐	☐	☐ closed

A.4 Celebrity Quotes

Experiment Questionnaire "Celebrity Quotes"

Group ________ Person __________

Indicate your preferred answer by marking an "X" in the appropriate box of the seven-point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply.

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

		FtF	**vCVE_im**	**sVC**	**vCVE_desk**
1	I could easily tell where my partner was looking.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
2	It involved a lot of effort to work together.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
3	I was always aware of my partner's presence.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
4	I always had a good sense of what my partner was doing.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
5	It was just like being face to face with my partner.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
6	It felt as if my partner and I were together in the same place.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
7	I could not contribute anything to the solution we came up with.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
8	My partner was "salient" to me during the task. *salient = noticeable: having a quality that thrusts itself into attention*	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
9	It was easy to understand my partner's intentions.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
10	I was often confused.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree
11	It was very easy to make myself understood.	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree	O O O O O O O strongly disagree · undecided · strongly agree

Comments: ____

		FtF	vCVE_im	sVC	vCVE_desk
12	I felt present in the virtual space.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree
13	I was completely captivated by the virtual world.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree
14	Somehow I felt that the virtual world surrounded me.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree
17	I had a sense of acting in the virtual space, rather than operating something from outside.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree
18	I was not very aware of my real environment while moving around in the virtual room.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree
19	There were moments when the virtual environment was the reality for me.		O O O O O O O strongly disagree undecided strongly agree		O O O O O O O strongly disagree undecided strongly agree

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

	FtF	vCVE_im	sVC	vCVE_desk
18	O O O O O O O warm cold	O O O O O O O warm cold	O O O O O O O warm cold	O O O O O O O warm cold
19	O O O O O O O impersonal personal	O O O O O O O impersonal personal	O O O O O O O impersonal personal	O O O O O O O impersonal personal
20	O O O O O O O sensitive insensitive	O O O O O O O sensitive insensitive	O O O O O O O sensitive insensitive	O O O O O O O sensitive insensitive
21	O O O O O O O unsociable sociable	O O O O O O O unsociable sociable	O O O O O O O unsociable sociable	O O O O O O O unsociable sociable
22	O O O O O O O active passive	O O O O O O O active passive	O O O O O O O active passive	O O O O O O O active passive
23	O O O O O O O pleasant unpleasant	O O O O O O O pleasant unpleasant	O O O O O O O pleasant unpleasant	O O O O O O O pleasant unpleasant
24	O O O O O O O formal spontaneous	O O O O O O O formal spontaneous	O O O O O O O formal spontaneous	O O O O O O O formal spontaneous
25	O O O O O O O positive negative	O O O O O O O positive negative	O O O O O O O positive negative	O O O O O O O positive negative

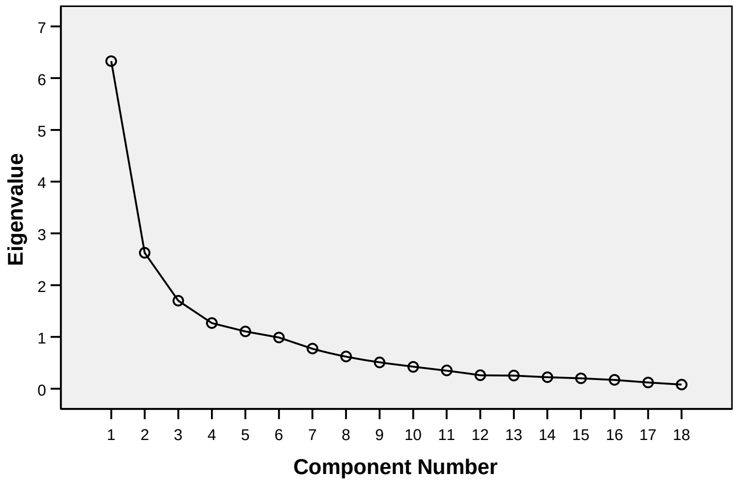
Comments: ____

Appendix B

Factor Analysis

B.1 Experiment "Desert Survival Game": Exploratory questionnaire items.

Scree Plot



Rotated Component Matrix(a)

	Component				
	1	2	3	4	5
There was a great sense of realism in the environment	.835				
I had a good feel of the people at the other end	.797				
It was like a face-to-face meeting	.917				
It felt as if we were all in the same room	.917				
The team's task performance was [poor...excellent]	.444	.514			
The task performance was highly affected by the communication medium. [disagree...agree]					.827
The overall flow of communication between the team was [poor...excellent]]	.577	.462			
I waited for my turn to speak. [never...often]			.887		
The others waited till it was their it was their turn to speak [never...often]			.874		
Turn taking was complicated [disagree...agree]		.604			
We were talking over one another [never...often]		.785			
I felt like i was interrupted [never...often]		.812			
I interrupted the others [never...often]		.800			
I was confused [never...often]	.304	.599		.456	
The others seemed to be confused [never...often]		.697		.484	
There was a lot of time when no one spoke at all [never...often]				.753	
I used body language to express myself [never...often]	.437		.416	.355	.487
The others used body language to express themselves [never...often]	.395		.463	.377	.468

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 10 iterations.

Appendix C

Transcription analysis

C.1 Transcription excerpt

Excerpts of a dialogue of two female participants, recorded during experiment "Celebrity Quotes" (condition vCVE_desk).

Person 1:	I am gonna go look at the pictures
Person 2:	Yeah. me too.
Person 1:	Oh, you are right here! Oh no! OK.
Person 2:	So now the quotes.
Person 1:	OK. How about you go stand at the quotes and read me quotes. And I look,... or I go read the quotes and you look at the people.
Person 2:	I am looking at the quotes. Oh, OK.
Person 1:	have a look at the people too and then we will have a good idea of what both is there.
Person 2:	Hmm. Well.
Person 1:	I am gonna go look at the people.
Person 2:	I think that Harry Potter said the thing about QUOTE.
Person 1:	OK. Do you want to put that in?
Person 2:	OK. I go.
Person 1:	So Harry Potter is number 46.
Person 2:	46?
Person 1:	yeah.
Person 2:	OK. and the quote is Quote_TAG.
Person 1:	OK. good.
Person 2:	OK.

...

. . .

Person 1+2: 1: I think...hello!....I think that
2: Hello! Hahaha.

Person 1: Mother Teresa said this one, QUOTE_TAG.
QUOTE.

Person 2: And so who would, who, hmm, what would the Buddha say?

Person 1: He would say to you that he is the Dalai Lama.

Person 2: It's not Buddha?

Person 1: The Dalai Lama, yeah. Hmm. OK.
We have to think about that one.

Person 1+2: 1: How about... Yeah. That's probably it.
2: QUOTE? Or...

Person 2: ..., or QUOTE?

Person 1: Yeah. That's true. There is too many of them. They are the same!
I think that, hmm, maybe QUOTE_TAG is Mike Tyson,
do you remember?

Person 2: Who?

Person 1: The boxer, with the tattoo on his face.

Person 2: Ah, yeah. Maybe.

Person 1: Because, do you remember the saying "Be Like Mike"?

Person 2: OK.

Person 1: But I don't know.

Person 2: I don't know either.

. . .

...

Person1: We still have the QUOTE_PART left.
Hmm, and QUOTE.

Person 2: Maybe it could be Marylin Manson?

Person 1+2: 1: Hmm...
2: Who said that?

Person 1: I have no idea. QUOTE. I don't know.
Do you wanna just not put the last three in?
Because if we get it wrong, we will loose a point.

Person 2: Yeah, yeah. You are right.

Person 1+2: 1: Doesn't... Hmm.
2: Yeah. Maybe we should stop.

Person 1: OK.

Person 2: Stop here.

Person 1: So do you want to click stop on that?

Person 2: OK. Stop!

C.2 LIWC2001 Output Variable Information

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, youre	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, youll	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
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

Table C.1: LIWC2001 test dimensions and examples