KTH Numerical Analysis
and Computer Science

The Effect of Modality on Social Presence, Presence and Performance in Collaborative Virtual Environments

EVA-LOTTA SALLNÄS

Doctoral Thesis
Stockholm, Sweden 2004

TRITA NA-0404
ISSN 0348-2952
ISRN KTH/NA/R--04/04--SE
ISBN 91-7283-707-1

Numerisk analys och datalogi
Kungliga Tekniska Högskolan
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av filosofie doktorsexamen 2004-03-24 kl 10.00 i Kollegiesalen, Kungliga Tekniska Högskolan, Valhallavägen 79, Stockholm.

Tryck: Universitetsservice US AB

ABSTRACT

Humans rely on all their senses when interacting with others in order to communicate and collaborate efficiently. In mediated interaction the communication channel is more or less constrained, and humans have to cope with the fact that they cannot get all the information that they get in face-to-face interaction. The particular concern in this thesis is how humans are affected by different multimodal interfaces when they are collaborating with another person in a shared virtual environment. One aspect considered is how different modalities affect social presence, i.e. people's ability to perceive the other person's intentions and emotions. Another aspect investigated is how different modalities affect people's notion of being present in a virtual environment that feels realistic and meaningful. Finally, this thesis attempts to understand how human behaviour and efficiency in task performance are affected when using different modalities for collaboration.

In the experiment presented in articles A and B, a shared virtual environment that provided touch feedback was used, making it possible to feel the shape, weight and softness of objects as well as collisions between objects and forces produced by another person. The effects of touch feedback on people's task performance, perceived social presence, perceived presence and perceived task performance were investigated in tasks where people manipulated objects together. Voice communication was possible during the collaboration. Touch feedback improved task performance significantly, making it both faster and more precise. People reported significantly higher levels of presence and perceived performance, but no difference was found in the perceived social presence between the visual only condition and the condition with touch feedback.

In article C an experiment is presented, where people performed a decision making task in a collaborative virtual environment (CVE) using avatar representations. They communicated either by text-chat, a telephone connection or a video conference system when collaborating in the CVE. Both perceived social presence and perceived presence were significantly lower in the CVE text-chat condition than in the CVE telephone and CVE video conference conditions. The number of words and the tempo in the dialogue as well as the task completion time differed significantly for persons that collaborated using CVE text-chat compared to those that used a telephone or a video conference in the CVE. The tempo in the dialogue was also found to be significantly higher when people communicated using a telephone compared to a video conference system in CVEs. In a follow-up experiment people performed the same task using a website instead, with no avatar but with the same information content as before. Subjects communicated either by telephone or a video conference

system. Results from the follow-up experiment showed that people that used a telephone completed tasks significantly faster than those that used a video conference system, and that the tempo in the dialogue was significantly higher in the web environments than in the CVEs.

Handing over objects is a common event during collaboration in face-to-face interaction. In the experiment presented in article D and E, the effects of providing touch feedback was investigated in a shared virtual environment in which subjects passed a series of cubic objects to each other and tapped them at target areas. Subjects could not communicate verbally during the experiment. The framework of Fitts' law was applied and it was hypothesized that object hand off constituted a collaboratively performed Fitts' law task, with target distance to target size ratio as a fundamental performance determinant. Results showed that task completion time indeed linearly increased with Fitts' index of difficulty, both with and without touch feedback. The error rate was significantly lower in the condition with touch feedback than in the condition with only visual feedback. It was also found that touch feedback significantly increased people's perceived presence, social presence and perceived performance in the virtual environment. The results presented in article A and E analysed together, suggest that when voice communication is provided the effect of touch feedback on social presence might be overshadowed. However, when verbal communication is not possible, touch proves to be important for social presence.

Acknowledgements

The journey I embarked on when starting this work has now come to an end as I am writing these words. It has been both the most exhilarating and agonizing period in my life. I would like to thank a whole bunch of kind and generous people who made this thesis possible.

First of all, I would like to express a very special gratitude to my supervisor Kerstin Severinson Eklundh for her unwavering support and confidence in me and my work. Her guidance has been of the greatest importance, and I was probably not even fully aware of her subtle influence on my path finding most of the time. Kerstin has always taken time answering my questions and has continuously encouraged me in my work.

My second supervisor, Ann Lantz has kept me firmly on the ground, helped me see problems in a wider perspective and she has frequently grounded me in our common background in psychology. Yngve Sundblad, my third supervisor, is the magician who makes dreams come true. Without his enthusiasm over my more elaborate ideas, in particular the haptic laboratory might not have been realized. Yngve and Ann also introduced me to the field of human computer interaction way back during my undergraduate studies.

I would like to thank Kirre and Calle for the most uncomplicated, fun and successful interdisciplinary cooperation I have ever experienced. The project about virtual environments that Anders made possible was also very interesting and rewarding. Thank you Anders for the interesting discussions we had. I would also like to thank Jin, the best Master thesis student one could ask for, whose work was important for my research, and who became a very dear

friend. Shumin has influenced and contributed a lot to my research. I thank him for being such an energetic, inspiring, friendly and open-minded person. I am very grateful for the generosity that many other people have shown me. In times when I needed advice regarding statistics, Bo always had time for me and he revealed the mysteries of statistical analysis with great patience. I would also like to thank everyone at AMT and especially Kjelle for his professional support. Thanks Maria Engström for making the picture on the cover and thank you Adam for the layout of this thesis.

All the colleagues at IPLab and CID have made the work environment there warm and stimulating. The discussions among the truly interdisciplinary mix of people have both confused and inspired me in my research, forcing me to constantly reflect on what I was doing. The long walks with Maria Normark, discussing every imaginable topic, have kept me sane. I am also grateful to all my other friends that have put up with my absent-mindedness; I will make it up to you now.

With all my heart to Johan, my travelling companion in life, for lending a shoulder to lean on when times get rough and for constantly surprising me in positive ways.

Mother and father, your unconditional love is the driving force in everything I do.

List of publications

This thesis is based on the following articles:

A. Sallnäs, E-L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction*, *7*(4), 461-476.

B. Sallnäs, E-L. (2001). Improved precision in mediated collaborative manipulation of objects by haptic force feedback. In G. Goos, J. Hartmanis and J. van Leeuwen (Series Eds.) and S. Brewster and R. Murray-Smith (Vol. Eds.), *Lecture Notes in Computer Science: Vol. 2058. Haptic Human-Computer Interaction* (pp. 69-75). Heidelberg, Germany: Springer.

C. Sallnäs, E-L. (conditionally accepted). Effects of communication mode on social presence, presence and performance in collaborative virtual environments. *Journal of Presence: Teleoperators and Virtual Environments*.

D. Sallnäs, E-L., & Zhai, S. (2003). Collaboration meets Fitts' law: Passing virtual objects with and without haptic force feedback. In M. Rauterberg, M. Menozzi & J. Wesson (Eds.), *Proceedings of INTERACT´03* (pp. 97-104). Amsterdam: IOS Press.

E. Sallnäs, E-L. (submitted). Passing virtual objects collaboratively with and without haptic feedback: Effects on social presence, virtual presence and perceived performance.

The articles will be referred to by the letters listed above.

Contents

"Social psychologists regard their discipline as an attempt to understand and explain how the thought, feeling, and behavior of individuals are influenced by the actual, imagined, or implied presence of others."

Allport, 1954 (p.3)

1 Introduction

People rely to an increasing extent on media for communication and collaboration, and as technology becomes more sophisticated, more human senses can be supported by media. It is now not only possible to hear and see but also to touch the same objects as a remote person and for example feel the other person's forcefulness when shaking hands. The purpose of this thesis is to explore how different media that support seeing the other, hearing the other or even support touching the same objects as someone else in a virtual environment, affect remote collaboration. The work in this thesis springs from a fascination for people's intuitive understanding of the benefits and limitations of a medium and for people's ability to cope with new technology and to use it in unexpected ways.

It is not relevant today, to make a distinction between digital and analogue media as these become more and more integrated. Metaphors are now mixed when computers and telecommunication technology converge into seamless new media. Virtual reality is an example of such a technology where many modalities can be used and where some information is digital whereas some can be analogue. This thesis examines how multimodality affects people's ability and experience of collaborating in virtual environments.

Human computer interaction studies investigating single user interfaces with touch feedback have only been conducted since the 1970's (Salisbury et al., 1995). The research about touch feedback in shared environments, that is the focus in the thesis, was almost nonexistent a few years ago. In a virtual environment with touch feedback, people can feel the shape of graphical objects, the weight, surface friction and softness of the object, the forces that

other people apply on objects as well as collisions between objects. Earlier research on the effects of different communication media on communication and remote collaboration is quite extensive, but experimental studies comparing communication media in shared virtual environments are not as abundant. As a consequence, the research in this thesis has been very interdisciplinary. Theory from social psychology, organisational psychology, virtual reality, perception psychology, human computer interaction (HCI) and from the area of computer supported cooperative work (CSCW) has been utilised in order to grasp the research topic.

One point of departure in this thesis has been the social presence theory that was developed in the research area of social psychology for evaluation of telecommunication media (Short et al., 1976). Social presence as defined in this theory, has not been studied to a large extent in shared virtual environments. An aim is therefore to examine how the framework of social presence theory can be applied in evaluating collaborative virtual environments. The thesis examines how hybrid media like virtual environments affect the experience a person has of social presence when collaborating with another person in a remote environment, and how varying levels of social richness in media make a difference for how collaboration takes place. Research comparing telecommunication media has suggested that the richer the media used, in terms of the extent verbal and nonverbal information is transmitted, the more social presence is experienced in human-human interaction. It has also been argued that people are aware of the varying levels of social richness in different media and therefore choose different media for different purposes (Daft and Lengel, 1986; Katz and Tushman, 1978; Rice, 1993; Short et al., 1976). One of the questions investigated in this thesis is how communication media like text-chat, telephone and video conference affect the experience of social presence in shared virtual environments. Furthermore, it is investigated whether adding touch feedback makes human-human interaction richer in the way social presence theory predicts.

A second issue that is investigated in this thesis, is to what extent a multimodal virtual environment makes people experience presence: "that they feel that they are there in the remote environment" rather than at their actual geographical location. The thesis here builds on the theory of presence, in the research area of virtual reality (Slater and Wilbur, 1997; Hendrix and Barfield, 1996; Lombard and Ditton, 1997; Freeman et al., 1998; Ijsselstein and Ridder, 1998). The unique focus in this thesis is on investigating presence in shared virtual environments and to systematically contrast this phenomenon with social presence when a variety of modalities are provided. In the research area of presence the interest lies in understanding which media characteristics that affect the extent that people perceive a virtual environment as realistic, natural and meaningful (Witmer and Singer, 1998). Furthermore, what characteristics

make users feel in control when interacting in a virtual environment and make them feel isolated from the real world as a result of being immersed in the virtual environment. All these issues might be affected by the modalities used in a collaborative virtual environment that make it possible to communicate with others in various ways and to manipulate objects with or without touch feedback. One aim of this thesis is therefore to investigate, how different communication media and touch feedback affect people's perceived presence in shared virtual environments.

A third aim in the thesis is to investigate how providing different communication media and touch feedback affects the behaviour when people perform tasks together in multimodal shared virtual environments. This is a perspective that builds on research in the areas of HCI and CSCW. In these areas of research the interest is in many cases to measure efficiency in task performance and also users' satisfaction with the process and/or the outcome. In the case of comparing ways of communication like video conference, telephone or text-chat, the focus is in this thesis on the amount of communication, duration of collaboration and the tempo in the dialogue. In the case of adding touch feedback to the shared virtual environment, the focus is on how efficiently people can coordinate their actions in order to accomplish their task and how precisely they can do their task compared to when they do not get touch feedback. The actual behaviour has consistently been compared to people's perceived task performance, in order to establish if there is a relation between actual performance and how people perceive that they accomplish collaboratively. The aim has been to examine how different communication media and touch feedback affect joint task accomplishment in shared virtual environments.

In conclusion, this thesis investigates collaboration in shared virtual environments and the effects of systematically comparing communication modes and of providing touch feedback, on social presence, presence, joint accomplishment of tasks and perceived performance.

1.1 Outline of the thesis

The theoretical background in the thesis covers a number of topics that are relevant for understanding why the research was performed, interpreting the results in the experiments, and evaluating the contribution of the thesis.

The initial part of the theoretical background, *chapter two*, is an introduction to the domain of collaborative virtual environments. What constitutes a group is discussed and also how collaboration and communication are defined in the thesis. The meaning of the concept multimodality is discussed and finally, the technological systems used in the studies are introduced.

In *chapter three*, the frameworks that the research in this thesis is based on are presented. The first framework introduces the idea that groups in an organisation differ in their need and use of communication media. It is argued that different media are appropriate for different tasks due to their difference in "richness" and thus groups that have specialised roles in an organisation use media that suit their purposes. The social presence theory is a framework where the idea that it is possible to measure this "richness" is advanced and here the construct social presence is presented. In the area of virtual reality it is argued that the characteristics of a medium can affect the extent that a person experiences presence, a feeling of "being in a virtual environment rather than at the physical location". Finally, in this part of the text the similarities and differences of these frameworks and the ways that they complement each other are discussed.

In *chapter four*, earlier research about the benefits and disadvantages of different communication media is presented and discussed in relation to the theories elaborated on in chapter three. In *chapter five*, the touch modality is defined and issues concerning the psychology of touch are discussed. The effects that have been found in earlier research of providing touch feedback in single user interfaces are presented as well as results about how touch feedback affects collaboration in virtual environments.

Chapter six goes through and summarizes the aims, methods and findings of the studies that constitute the empirical part of the thesis. Conclusions are drawn in *chapter seven*, about the role played by different modalities supported in the experiments regarding the experience of social presence and presence, joint task accomplishment and perceived task performance. Finally some methodological issues are discussed and future research is suggested.

The last part of the thesis, consists of the papers that describe the empirical studies performed in this thesis. They are identified by the letters A-E.

The study presented in *article A* was performed in cooperation with Kirsten Rassmus-Gröhn and Calle Sjöström from the Division of Rehabilitation Engineering Research (CERTEC), Lund Institute of Technology (LTH). The purpose of that study was to determine how a remote collaboration situation would be affected if people could use the sense of touch when solving manipulation tasks in a shared virtual environment. The variables measured were completion time, perceived performance, social presence and virtual presence. I planned and conducted the experiment and Kirsten and Calle programmed the application and managed the technology used for the experiment. In *article B* a further analysis, performed by me, of the experimental data collected in the study described above concerning precision in task performance was presented.

In *article C* a study is presented where the purpose was to investigate if pairs of people behaved differently or if their experience of the interaction varied depending on if they communicated by text-chat, a telephone connection or a video conference when they completed a task together in a shared virtual environment. These results were also compared to performing the task in a web environment. The variables measured were completion time, the amount of words and the tempo in the dialogue, perceived performance, perceived social presence and perceived presence.

The research performed in *article D* was conducted in cooperation with Shumin Zhai from IBM Almaden Research Center who was a guest lecturer for one year at the Royal Institute of Technology. One aim was to examine if people that passed objects to each other in a shared virtual environment would benefit from getting touch feedback. A second aim was to test if Fitts' law held for a collaboratively performed tapping task. The time it took to pass an object and the error rate were measured. The work was divided in such a way that Shumin and I planned and designed the experiment together, I conducted the experiment and data collection and finally we analysed the results together. Questionnaire data obtained in this study was later analysed by me and the results were reported in *article E*. The questionnaires measured perceived performance, perceived social presence and perceived virtual presence.

2 Collaboration in multimodal virtual environments

Group communication and collaboration mediated by multimodal systems is a challenging and complex area of research. To master the research questions in this research area the perspective has to be truly interdisciplinary. The dynamics of groups are affected by many internal and external variables and the technology that mediates group interaction is one of those external variables.

2.1 Groups, collaboration and communication

Communication as a concept is used and defined in a lot of different ways with focus on aspects such as sender and receiver of messages, encoding-decoding, attribution of meaning, evoking of response, conveying meaning and so on. A social psychological definition of communication is formulated as follows (Cherry, 1957):

"the psychological signals whereby one individual can influence the behaviour of another "

This definition implies that constraints on these psychological signals can affect the possibility to influence another person. Technology that mediates

communication can for example be such a constraint. There are different levels of communication, and in this thesis the focus is on small-group (dyadic) communication. One definition of what a group is has been formulated by Schein (1965):

"a psychological group is any number of people who 1) interact with one another, 2) are psychologically aware of one another, and 3) perceive themselves to be a group"

Groups can be formal or informal and the communication in formal groups can also be either formal or informal. The criteria of the formal group are usually known for example by organisational structure whereas the criteria of the informal group are often not stated. Formal communication channels are those emanating from official sources and carrying official sanctions, and formal messages usually flow trough these channels, thus acquiring legitimacy and authenticity (Smith et al., 1972). Informal communication channels are not specified rationally, they develop through accidents of spatial arrangement, through friendships and through the varying levels of ability in the organisation's boxes. Informal communication has been identified as very important for groups in order for them to successfully accomplish their work (Kraut et al., 1993).

Theories about groups often categorise the functions and activities that groups do in order to fulfil the goal of the group and the goal of the organisation. It is interesting that a common denominator in many theories about groups is the emphasis put on the importance of both task and production on the one hand and maintaining and managing social relations on the other, in order for the group to attain its goals. According to the task circumplex model a formal group has three functions that are to produce, give members support, and maintain the group (McGrath, 1993). In another theoretical approach, researchers suggest that groups work concurrently on three interlocking activity tracks: task-process activities (manage tasks), relational activities (manage interpersonal relations) and topical focus (specific issues of concern to the group) (Poole et al., 1996). Similarly, Bales (1950) argues that in order to maintain the group, people engage in two general categories of communication: task-oriented and socio-emotionally oriented communication. This dual focus is also reflected in this thesis, where both the experience of the social richness in mediated interaction on the one hand, and the behavioural outcome of the collaboration on the other hand are taken into consideration. Finally, the extent that people are immersed in a mediated environment and *feel that they are present there* is evaluated.

Decision-making in groups through information exchange is affected by a lot of different facilitating and constraining factors. Relevant for the focus

of this thesis are the ways that technology can facilitate and constrain the information exchange in groups. The fit between technology features and the group task is one example. Here the theory of media richness (Daft and Lengel, 1986), argues that media can be arrayed along a continuum from rich to lean where more ambiguous and equivocal tasks should be performed in richer media. An integration of the media richness model with the task circumplex model has been suggested (McGrath and Hollingshead, 1994). When these two models are integrated it is argued that generating ideas and choosing a correct alternative could be done in a lean medium, whereas making consensus decisions or negotiating conflicts of interest should be done in a rich medium. Technology use is dynamic and feedback from the exchange of information, as it aggregates over time in groups, is affecting the technology used in a group. If technology does not meet the demands or is inefficient for the intended purpose it is soon exchanged to a more appropriate alternative.

Theory regarding collaboration addresses some essential research questions. How can general goals be divided into subgoals and activities? How can activities be distributed among groups or individuals? How can resources be allocated to different actors? How can information be distributed between different actors with the purpose of reaching general goals? Bannon and Schmidt (1989) suggest that cooperation is work processes, interrelated by content like for example to produce a product or service. In contrast to the notion that cooperation is bounded by a formal organisation, a cooperation process can involve actors from different organisations, on different locations. Cooperation could be guided both indirectly through the changing stages in a developing process, or directly by interpersonal communication. Cooperation can be done by a co-located group or by distributed semi-autonomous individuals that coordinate their work strategy themselves.

In order to cooperate, groups and individuals have to coordinate their activity through communication. Good communication is especially important in complex problem solving situations. The components that are important for coordination are activities, goals of activities, actors and interdependencies. Interdependencies are factors that need to be coordinated, because they are factors in an organisation that can go wrong and that can hinder accomplishment of goals. Interdependencies that are of importance are shared resources, to synchronise simultaneous activities and to manage processes where one activity has to be finished before the next activity can start. Underlying processes to coordination are decision-making, communication and perception of common objects like physical objects or shared databases. Malone and Crowston (1990) define coordination as follows:

"*The act of managing interdependencies between activities performed to achieve a goal.*"

Groupware are systems that support distributed or co-located cooperation, coordination and communication, and this concept is used in the area of computer supported cooperative work (CSCW). Groupware supports a number of functions that can be classified in a schema based on the way they satisfy needs regarding communication, coordination and cooperation in groups (Sauter et al., 1995). A frequent way of classifying these systems is in terms of their ability to support groups whose physical proximity varies along two continua of time and space (McLeod, 1996). In the time continuum there is groupware that supports synchronous and asynchronous work and in the space continuum groupware can support different degrees of distributed or co-located work.

Important aspects of social interaction that have to be considered when developing systems that support distributed communication and collaboration are: proximity and orientation, physical contact in formal situations, technological/social immediacy, physical appearance, non-verbal signals from the trunk and arms, facial signals, direction of eye-gaze, mutual gaze and intimacy, specific qualities of verbal versus visual communication (Short et al., 1976). These are aspects that people manage quite effortlessly in face-to-face situations in real life but that can be hard to adjust in a flexible manner in remote interaction. If for example the video image of a person is very small, that person can be perceived to be farther away than required compared to the interpersonal distance that would have been chosen between people for that particular situation.

2.2 Collaborative multimodal virtual environments

The concept of multimodality is used in a number of rather different ways. In some cases technology is categorised as multimodal because people can interact with it using output like for example gestures or verbal commands. In other cases the focus is on the senses that humans can use in order to get input from a system and from another human. The difference between multimodal systems and multi-media systems is that multimodal systems support more human modalities, which increases the bandwidth for information intake. In multi-media systems on the other hand, information is presented in many different ways but humans sometimes only use one or a few senses in order to perceive it. The focus in this thesis is mainly on adding modalities so that people can use more of their senses which increases the information bandwidth. However, it is not always easy to draw a line between multimedia and multimodality.

The five senses vision, hearing, touch, smell and taste help humans make sense of the world and grasp what constitutes self. Humans rely very much on the vision and hearing modalities for perception but the touch, smell and taste

modalities are also vital for understanding reality, for performing actions in our environment and for interacting with others. It has been argued that sensory modalities are specialised for different tasks, and that this specialisation emerges more strongly as the complexity of a task increases (Freides, 1974). Vision is generally dominant over both touch and audition for the perception of spatial location and vision is more effective than touch for perception of shape (Heller and Schiff, 1991). Evidence for dominance of audition (hearing) over vision has been found in the perception of temporal information such as judgment of sequence rate and perception of duration (Repp and Penel, 2002). Touch is at least as accurate as vision in the perception of texture and if vision is blurred people rely more on touch for perception of form (Heller and Schiff, 1991). It has been argued that the specific characteristics of any particular perceptual task should be considered, in relation to the specific properties of the sensory modality (modalities), that provide information for performance of the task (Heller and Schiff, 1991). However, perception is not independent of cognition. Memories, expectations, emotions and attitudes shape our perception, sometimes fooling us with illusions and sometimes helping us fill in incomplete information in order for us to perceive things correctly.

Virtual reality can be described as a technology that enables a user to experience and interact with a three dimensional computer generated environment (Axelsson, 2002). There are many kinds of virtual reality systems and a review of all the different technologies used in these systems is not possible in this thesis but the interested reader can find more information in Jää-Aro (2004) and Schroeder (2002). A virtual reality system can be more or less immersive. In some virtual reality systems the virtual environment is projected on several walls, the floor and the ceiling in a room that the user can stand and walk around in. The user then experiences that the virtual world totally surrounds her. In another kind of virtual reality system the virtual environment is projected in a helmet (head-mounted display systems, HMDs) so that the user that wears it can see the virtual environment inside it. Finally, the virtual environment can be displayed to the user at a personal desktop computer.

When several computers are linked together several users can simultaneously view and interact with this kind of virtual reality environment and they can interact and collaborate with each other in it. Such environments are referred to as shared virtual environments or collaborative virtual environments (CVE). A user usually has a graphical image, called an avatar, that represents them in the virtual environment and that can be seen by other users in the same environment. With the avatar, the user can navigate in the world by the keys on the keyboard or by a joystick or with other devices. A large number of different shared virtual environments have been developed. Some of these environments can be accessed on the internet and they are usually virtual places where people

socialize informally. The more expensive virtual reality systems are often used in research laboratories and are not available to most people.

In order to get touch feedback in a virtual environment a haptic or a tactile device can be used. In a shared haptic object-space people can coordinate joint movement of objects by signalling direction through haptic force and give objects to each other almost without verbal communication. Many of the touch displays that have been developed in recent years use one-point haptic interaction with the virtual world like the desktop Phantom (Figure 1) used in the studies presented in article A-B and D-E in this thesis, marketed by SensAble Technologies. The effect is somewhat like tracing the outline of an

Figure 1. One person holding a desktop Phantom pen in order to feel and grasp a graphical object and in this case pass it to another person that uses an identical system in another room. The Reachin Display system has a mirror that the user looks at instead of the screen. In this way the person can hold the Phantom pen under the mirror, which creates an experience of touching the object where you see it. Stereo vision is provided through Stereographics CrystalEyes 3 shutter glasses, which makes the image three-dimensional and people can thereby see the depth in the image.

object with your index finger in a thimble or holding a pen and recognising it through this information alone.

The only skin receptors affected by the display are those that are in contact with the pen or thimble. The haptic information is not primarily intended for the skin receptors of the human tactile system. However, it is impossible to separate the systems completely. The skin receptors provide pressure and vibration information present also in a haptic system. But it is the movement, the involvement of the kinesthetic and proprioceptic system, that provides the information necessary to the perception of the model as an object. Haptic sensing has been called the active touch because movement is needed for perception and it is the kind of touch we experience in everyday life. Haptic sensing is defined as follows (Appelle, 1991):

"the use of motor behaviours in combination with touch to identify objects"

Tracing the outline of a virtual object will eventually give the user some notion of the shape of the object. A distinction is usually made between haptic and tactile interfaces. The tactile interface is an interface that provides information more specifically for the skin receptors, and thus does not necessarily require movement of the hands or limbs. An example of a tactile device is the braille display.

In the following, the shared virtual environments used in the studies in the thesis will be described.

In the study presented in article C, people collaborated in a virtual exhibition that was developed in Active Worlds (Figure 2) from ActiveWorlds Inc. The virtual exhibition was displayed to the users using two lap-top computers. The

Figure 2. One user's view of another person and of information about different car models that are displayed in a shared virtual exhibition. The blue signs in the exhibition can be selected by the users and a video image is then shown to the right and they can in that way look at and listen to a movie clip that presents information about a car model. Users can navigate around with their avatars in the virtual exhibition, investigate the information and discuss it at the same time.

virtual exhibition consisted of an area with walls all around, and an exit in the middle of one of the walls. Information points were displayed in the exhibition, in the form of pictures and links to video clips in which information about the car models was presented. Both users were represented by an image of a person, an avatar. Each user could see the other person's avatar but not their own. They could move around in the world using the keyboard and they could position themselves in relation to the information and also in relation to the other person. In this way it was possible for them to see in which direction the other person was looking and where the other person was going in the virtual environment.

In the experiment presented in article C, three different communication media were tested in combination with the shared virtual exhibition. In one condition, users had a large video conference image 30 degrees on their left hand side and the PowerBook with the virtual exhibition directly in front of them (Figure 3). They also had headsets and talked to each other via a telephone connection. In a second condition users only had a telephone connection and two headsets. In a third condition users communicated via text-chat which is a feature in Active Worlds.

Figure 3. Two people collaborating using a video conference a telephone connection and headsets. In this picture the web-environment was used instead of the virtual environment in Active Worlds.

Haptic feedback systems now make it possible not only to use vision and hearing but also to use the touch modality when interacting in virtual environments. A virtual environment with haptic feedback was developed for the experiment presented in article A and B that consisted of a three dimensional room with eight cubes and where users were represented by avatars in the shape of one blue and one green sphere (Figure 4).

Figure 4. A haptic virtual environment with two users lifting one cube together, feeling the forces from the other person.

In this virtual environment users could feel the force resistance from the walls, floor and ceiling as if they had put their hand into a physical box. They could feel the shape and weight of the cubes and the forces that the other person applied at objects or at the other person's avatar. They could in a simplified way grasp the other person and symbolically shake hands. It was not possible to grasp the cubes in this environment, so in order to lift them users had to push from each side of a cube.

In the experiment presented in article D and E, another virtual environment with touch feedback was developed that consisted of a room, four shelves, six cubes and two avatars in the form of one green and one blue sphere (Figure 5). Also in this environment users could feel everything that was visible in the virtual environment.

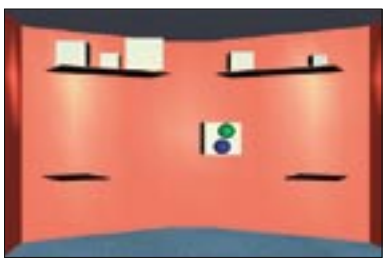

Figure 5. Two users passing a cube from one person to the other in a virtual environment with touch feedback.

In this environment it was possible to grasp a cube by pushing a button on the interaction device. Consequently users could both pull at the object and at the same time feel the other person's pulling actions through the object. It was in this study examined if this would facilitate passing cubes from one person to the other.

3 Research perspectives

In organisational and social psychological studies of mediated interaction, the interest is in the capacity of media to transmit social information. The focus has mainly been on different communications media like e-mail, telephone or videoconference compared to face-to-face interaction. A general argument is that the richness of the communication medium (Katz and Tushman, 1978; Daft and Lengel, 1986; Rice, 1993) or the level of perceived social presence (Short et al., 1976), has to be matched with the task in order for collaborators to accomplish tasks satisfactorily. In virtual reality research one theoretical perspective focuses on the extent to which a person has a feeling of being present in virtual reality environments, or in other media such as films or books (Hendrix and Barfield, 1996; Lombard and Ditton, 1997; Slater and Wilbur, 1997). The thesis has been motivated to a large extent by these theories that come from different disciplines. These theoretical perspectives all claim that supporting more human modalities make the interaction more efficient or richer in different ways. The aim in this thesis has been to consistently apply the theories in the studies and to compare the results derived in order to add knowledge that might develop the theories further.

3.1 The organisational perspective

The contingency theory was first presented in an explicit way by Lawrence and Lorsch (1967) and has been very influential for the research and theorising on communication in organisations (Katz and Tushman, 1978; Daft and Lengel, 1986; Kraut et al., 1992; Burrel and Morgan, 1993; Rice, 1993).

The contingency theory asserts that the effective operation of an enterprise is dependent upon there being an appropriate match between its internal organisation and the nature of the demands placed upon it by its tasks, its environment and the needs of its members. This challenged the view of classical management theories that sought to specify universal principles of organisation and management without taking the specificities of the demands placed on different organisations into consideration (Burrel and Morgan, 1993). In the view of contingency theory when applied to communication in organisations, the more people are dealing with complex tasks, the greater is the work related uncertainty and the greater are the communication requirements (Katz and Tushman, 1978). It had been found that different types of tasks required different amounts of information and that task complexity was associated with greater amounts of communication (Hage et al., 1971; Van de Ven et al., 1976). Katz and Tushman (1978) argue that trying to foster more communication in a general way in organisations might not be useful. They showed that neither increasing internal nor external communication in a general way improved performance per se for any of the project areas that they studied. The reason for this is that the benefit of communication was related to what kind of work different groups performed in the organisation, which means that it is more important to solidify communication patterns that are relevant for each specific type of work.

High performing research groups with complex and more universally defined projects, communicated a lot within their own projects and with other areas within the laboratory. The external contacts that research groups had were mainly with external professional areas. With respect to both outside professional contacts as well as internal contacts in the research groups all group members had widespread face-to-face contacts with other areas.

High performing developmental and technical service projects which performed more routine tasks, communicated mostly with the other corporate areas. These groups had most external contacts with customers, external consultants and suppliers. In the developmental and technical service projects specialized boundary roles, as for example a group leader, were needed to deal effectively with external contacts indicating that these groups relied more on a hierarchical structure. The results from this study by Katz and Tushman (1978) are consistent with the ideas of Lawrence and Lorsch (1967) and Galbraith (1973), that integrating mechanisms for high performing units are contingent on the nature of the unit's work.

In media richness theory, Daft (1998) builds on the contingency theory and focuses on the capacity of different communications media to support rich communication. Information richness is defined as "the information carrying capacity of data". In media richness theory it is argued that a good match

between communication media and organisational tasks maximises efficiency and satisfaction (Daft and Lengel, 1986). In this approach the central concepts are information ambiguity and media richness. Information ambiguity means that issues cannot be objectively analysed and understood, and that additional data cannot be gathered that will resolve the issue. The idea of ambiguity is probably inspired by the work of Katz and Tuchman (1978) regarding communication patterns in what was called complex and universally defined projects, that was described earlier in this chapter. In media richness theory a difference is made between an ambiguous and an uncertain situation. In contrast to how an ambiguous situation can be handled, a large amount of information can resolve an uncertain situation, which thus can be managed by collecting more facts in for example databases. In order to cope with an ambiguous situation on the other hand, to understand ill-defined issues, managers need to process richer information, by for example discussing face-to-face. So, an uncertain situation is complex in the way that a lot of more information is needed in order to solve it, whereas an ambiguous situation is complex in the way that a rich medium or even face-to face interaction is needed in order to solve it. Discussions face-to-face make it easier to understand emotional aspects of a situation and make it easier to reach an agreement that gives guidance on how to respond to the situation.

The communication channels used in organisations were organized by Daft (1998) into four categories ranging from highest to lowest in richness; face-to-face, telephone and voice-mail, written addressed documents like letters and faxes and finally written impersonally addressed documents. Face-to-face was considered the richest medium as it provides many non-verbal cues and was considered the best channel for mitigating ambiguity. Telephone and voice-mail were the second richest media as feedback is fast and messages are personally focused. Written addressed documents were considered lower in richness because feedback is slow and personal cues are minimal. Finally, written impersonally addressed documents like bulletins, computer bases and printouts, are the leanest channel because feedback is not possible and they are not personal. Channels low in richness are considered lean because they are effective for conveying a large amount of data to numerous people.

Kraut et al., (1992) argue that it is important to make it clear that there is a difference between the level of interactivity and the level of expressiveness when talking about rich and impoverished communication channels. Kraut et al. (1992) build on Clark and Wilkes-Gibbs (1986) work and argue that a medium that provides interactivity permits communication partners to exchange information rapidly, adjusting their messages in response to signals of understanding or misunderstanding and to interruptions or questions. The term expressiveness on the other hand has similarities with the term richness in the media richness theory (Daft and Lengel, 1984; Daft and Lengel, 1986)

and the concept of social presence (Short et al., 1976). A medium that provides expressiveness allows individuals to convey not only the content of their ideas but also intensity and subtleties of meaning through intonation, facial expression, or gestures.

Media richness theory suggests that not only is a larger amount of communication needed in complex situations, as explained by contingency theory, but the richness of the channel for communication needs to be congruent with the level of ambiguity of the task. However, media richness theory does not explain more explicitly what the medium characteristics are that affect the richness of a medium or how different characteristics of a medium affect the users' experience of richness. In fact, the definition of information richness as formulated in media richness theory is not very satisfying if the aim is to design or evaluate communication systems. Issues regarding what constitutes richness have to a larger extent been the focus of research in the area of social psychology of communication.

3.2 The social psychological perspective

In the social psychological perspective, the functions of verbal and nonverbal communication are very much the focus of investigation. The unit of analysis is small groups (often dyads) of people rather than organisations. The symbolic interactionist George Herbert Mead emphasised the importance of both verbal and nonverbal communication for understanding social phenomena according to Burrel and Morgan (1993).

In this thesis the theory of social presence has been of great importance. Social presence theory evolved through research about efficiency and satisfaction in the use of different telecommunications media. Social presence refers to the feeling of being socially present with another person at a remote location. Short, Williams and Christie (1976) introduced and defined the term social presence as follows:

"the salience of the other in a mediated communication and the consequent salience of their interpersonal interactions"

The construct social presence can be traced back to Mehrabian´s (1969) concept of immediacy that he defined as "those communication behaviours that enhance closeness to and nonverbal interaction with another". Immediacy is a measure of the psychological distance that a person puts between her/himself and another person. This distance can be shown through speech that reflects either feelings of closeness or of separation. Heilbronn and Libby (1973) mean that other factors like the medium a person chooses to use for

communication, for example text or voice, can convey immediacy or non-immediacy. If a person sends an email the information might be perceived as having different connotations compared to if a telephone was used. In general, the more information a medium can transmit, the greater the immediacy.

Social presence is conceived to be a subjective quality of a medium and is assumed to vary between different media which affects the nature of the interaction (Short et al. (1976). Social presence is regarded as a single dimension that represents a cognitive synthesis of several factors that are naturally occurring in face-to-face communication. Among these are the capacity to transmit information verbally but also by nonverbal cues like gestures, facial expression, direction of gaze and posture. The inability of media to transmit nonverbal cues can have negative effects on interpersonal communication. The medium's capacity to transmit this kind of cues affects, according to social presence theory, the level of social presence, that is the extent to which a medium is perceived as sociable, warm, sensitive, personal or intimate when it is used to interact with other people.

Short et al. (1976) argued that the approach of measuring efficiency through for example measures of task completion time and amount of communication used in different media is unsatisfying for a number of reasons. It is not always unambiguous what task completion time means, or how to interpret amount of communication in terms of efficiency. It has often been found that people complete a task faster in a telephone setting which might mean that the communication is more task focused, formal and efficient. But if the task is of an ambiguous character, video might improve the quality of the task outcome because people have longer dialogues and more socioemotionally oriented communication. The fact that task completion time is often longer on the other hand, in a text communication setting than if audio is provided might in fact just mean that it is cumbersome to write text. Because of these limitations of the efficiency approach it is important to take the communication actors' feelings into account and this is the reason why Short et al. developed the Social presence theory. Social presence is a subjective quality of a medium not an objective quality even though the subjective perception is dependent on the objective characteristics of a medium.

Short et al. (1976) refers to Dashiell's experiment that was reported in 1935 in which it was found that it was not the objective fact that there were others in a nearby room that affected subjects performance but the subjective fact that subjects were aware of this or not. This is called the social facilitation/inhibition effect (Allport, 1954) which means that people's performance of a simple task is facilitated by the actual presence or conviction that others are present nearby, but is inhibited by this in the performance of a complex task. The phenomenological essence of looking at social presence as a subjective

quality of a medium is that it is a person's perception of a communication situation that counts and that is in many ways a much more reliable and valid measure as it takes into account the whole complexity of variables that are specific for each communication event. Following from this is that if we know the level of social presence it is more straightforward to make prediction of when a medium will be used. This is especially useful because of the fact that people are aware of what level of social presence is efficient for what tasks.

The social presence theory relates to yet another concept in social psychology namely Argyle and Dean's (1965) concept of intimacy. People are both attracted to and repelled by other people in a communication situation and therefore strive for equilibrium of intimacy through for example seating position (the more intimate people are the closer they sit). A number of factors affect the level of intimacy like physical proximity, eye contact, amount of smiling and intimacy of topics of conversation.

$$\text{Intimacy} = f \begin{cases} \text{eye contact} \\ \text{physical proximity} \\ \text{intimacy of topic} \\ \text{amount of smiling} \\ \text{etc.} \end{cases}$$

Argyle and Dean (1965) cite George Simmel (1921) who described one of the factors, namely eye contact as, "a wholly new and unique union of two people" and who remarked that it "represents the most perfect reciprocity in the entire field of human relationship". The idea of equilibrium of intimacy was inspired by research that had shown the different functions that eye contact as well as physical proximity had during human-human communication. Feedback from looking at the other's face, and especially in the region of the eyes convey information of how a message has been received. There is more eye contact if two persons like each other and if they are cooperating rather than competing (Goffman, 1963). There is less eye contact if there is tension in a relationship between people. Eye contact regulates turn taking in a conversation and a person gazes more when she is listening than when she is speaking (Exline et al., 1965). People differ in the amount of eye contact they need and tolerate in order to feel appreciated or threatened respectively. People also use eye contact to regulate how much feeling they reveal and to discover others' intentions. It was for example shown that people gaze less at each other if they have been induced to cheat. Finally, eye-gaze is used in order to establish the relationship between people like friendship or hate and dominance or submission. Argyle and Dean (1965) propose that eye contact, physical proximity and other social components affect the level of intimacy and that people strive for a

state of equilibrium. If this equilibrium is disturbed in one of its constituent components there will be compensatory changes for the other components. However, if it is not possible to compensate because the components are fixed, and the deviation is too extreme, people will either feel to close or feel deprived of affiliative satisfactions. Short et al. (1976) argue that social presence should be included as a component in the equilibrium of intimacy model.

Social presence theory does not elaborate much on the use of different communication technologies at an organisational level. Short et al. (1976) do however, investigate the effects of different communication media compared to face-to-face communication on the outcome of meetings in relation to the type of activities performed in these meetings. Based on a number of studies it is concluded that the outcome of a meeting can be significantly affected by whether the meeting is face-to-face or if voice only or video/voice systems are used. Contrary to expectations the overall effect obtained of medium was not as large as hypothesised. Face-to-face meetings showed several significant differences from mediated meetings. However, only in some cases a significant difference on the outcome of meetings was found when voice only and video/voice systems were compared. Meetings involving attitude change or negotiation or involving getting to know strangers were found to be more sensitive than other situations regarding which communication medium was used. Yet in many situations, such as problem-solving interactions, the omission of non-verbal cues had little effect on the outcome of the task. Short et al. suggest that non-verbal communication only has effects on the outcomes of mediated interactions to the extent that it determines feelings of social presence and to the extent that the activities performed are sensitive to the "apparent distance" between people.

Short et al. (1976) argue that people are more or less aware of the degree of social presence of a medium, and choose to use a medium that they perceive to be appropriate for a given task or purpose. The order of media that was suggested by Short et al. (1976), based on the results from their studies, regarding the level of social presence was from higher to lower: face-to-face, television (video), multispeaker audio, telephone, and finally business letters.

Rice (1993) builds on the work of Daft et al. (1987) and Short et al. (1976) and argues that those researchers apparently independent of each other formulated similar hypotheses regarding media appropriateness. Rice (1993) argues that the essential underlying principle in both theories is that a good match between the characteristics of a medium (such as high in social presence or media richness) and one's communication activities (such as socioemotional activities like getting to know someone, or equivocal tasks like strategic decision making) will lead to better (more effective, satisfying, etc.) performance. Both theories emphasise that communication media differ in how they "(a) can overcome

various communication constraints of time, location permanence, distribution and distance, (b) transmit the social symbolic and non-verbal cues of human communication and convey equivocal information" (Rice, 1993). The focus in Rice's research was on perceived appropriateness and in a number of studies that he conducted people ranked how appropriate they perceived different media to be for different communication activities through a questionnaire. Social presence scales were used in order to measure perceived appropriateness regarding different media, rather than media richness scales, as Rice (1993) argues that the latter have been much less empirically evaluated.

An important conclusion that Short et al. (1976) make is that the phenomenon of social presence takes more aspects into consideration than the verbal and non-verbal information in itself. Social presence is in a way a phenomenological variable that includes people's perception of being together at the same place but also the extent that they are aware of the fact that they are separated. The extent that persons get so immersed in a mediated "virtual" environment that they forget or become unaware of the fact that they are separated probably makes a large difference in itself. This leads us to the next theoretical perspective which concerns presence in virtual environments.

3.3 The virtual reality perspective

In the area of virtual reality, one of the aims is to generate an experience of being in a computer-generated environment that feels like reality. The importance of the extent to which a person has a feeling of being present in virtual reality environments, or other media as films or books, is therefore increasingly recognised (Hendrix and Barfield, 1996; Lombard and Ditton, 1997; Slater and Wilbur, 1997; Freeman et al. 1998; Ijsselstein and Ridder, 1998). Presence has been defined by Witmer and Singer (1998) as follows:

"*the subjective experience of being in one place or environment, even when one is physically situated in another*"

Applied to teleoperations, Witmer and Singer (1998) define presence as the sensation of being at the remote work site rather than at the operator's control station.

Presence has also been defined as "a state of consciousness, the psychological state of being there" (Hendrix and Barfield, 1996; Slater and Wilbur, 1997). In some research the aim is that a person should feel physically present in the media environment to the extent that the person should not be able to distinguish between the real and the mediated world (Schloerb, 1995). Thus a person is subjectively present in some particular environment if the person

perceives that he or she is physically present in that environment. Minsky (1980) defines telepresence as "feeling like you are actually there at the remote site of operation", while he defines virtual presence as "feeling like you are present in the environment generated by the computer".

These definitions seem bipolar, as if a person either is experiencing full presence or is not experiencing presence at all. Witmer and Singer (1998) argue however that one can compare this concept with attention, or with the concept of focus, which humans experience in varying degrees and more importantly selectively. Selective attention is the tendency to focus on selected information that is meaningful and of particular interest to the individual (Fontaine, 1992). Research has shown that attention is guided by the meaningfulness of the information presented. From this follows that the information in the media environment must be meaningful, to maintain the individual's focus.

Other psychological concepts that are of interest in relation to presence are involvement and immersion. Involvement is a psychological state experienced as a consequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related activities and events (Witmer and Singer, 1998). Accordingly, in order to maintain the individual's focus and sense of presence, the information in the media environment must be meaningful to that individual. Involvement depends on the degree of significance or meaning that the individuals attach to the stimuli, activities, or events. Furthermore, as users focus more attention on the virtual reality stimuli, they become more involved in the virtual reality experience, which leads to an increased sense of presence. In order to maintain a high level of involvement, distracting events in the physical locale must be limited, or it must be possible for the individual to integrate them in the virtual environment in a meaningful way. A higher level of involvement increases a person's feeling of presence.

Another aspect, immersion, also affects the level of perceived presence. Immersion is a psychological state characterised by perceiving oneself to be enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences (Witmer and Singer, 1998). Factors that affect immersion include isolation from the physical environment, perception of self-inclusion in the virtual environment, natural modes of interaction, and control and perception of self-movement. A virtual world that produces a greater sense of immersion increases the level of presence.

Slater and Wilbur (1997) instead regard immersion as a feature of a technology, defined as "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant". Inclusive means the extent to which physical reality is shut out. Extensive indicates the range of sensory modalities accommodated.

Surrounding is the extent to which the virtual environment is panoramic rather than limited to a narrow field. By vivid they mean that the resolution and the quality of the displays can differ which affects the level of immersion. Two more factors are of importance for the level of immersion according to Slater and Wilbur (1997). One of them is how good the match is between the users proprioceptive feedback about body movements and the information generated on the displays of the users virtual body movements. This means that there should be a good match between the perceived movement in the virtual world and the actual movements that the person makes in the real world. Finally, the content in the virtual environment needs to be realistic if it represents a particular context and there should be a story-line that is comprehensible, meaningful and self-contained. The extent to which the user can influence the content and change the unfolding of events and the virtual world affects the level of immersion.

A person's experience of presence in a virtual environment has been measured by several dimensions in questionnaires. Examples of such dimensions are control factors, sensory factors, distraction factors and realism factors. Witmer and Singer (1998) have made a detailed review of the dimensions that several researchers find essential to include in a questionnaire that measure presence.

Social aspects concerning virtual environments are starting to become recognized in the research area that focuses on perceived presence in virtual reality. The reason is probably that an increasing number of people socialize in virtual places, collaborate in professional simulations or compete in multi-user games. In virtual reality research, the concepts of togetherness and co-presence are used in order to address issues of social interaction (Durlach and Slater, 2000; Slater, et al. 2000; Slater and Steed, 2002; Mc Laughlin et al., 2003; Biocca et al., 2001). However, I argue that these concepts in many cases focus more on the extent that people perceive that an avatar is representing a human being in a realistic way, than on the extent that the interaction is perceived as more or less rich. The notion of togetherness has been investigated as "a sense of being together in a shared virtual space" (Durlach and Slater, 2000). This notion of togetherness is argued to be a counterpart to the notion of one person feeling as if she is present in a virtual environment.

Heeter (1992) divides the concept presence in three dimensions; personal presence, social presence and environmental presence. Personal presence is according to Heeter a measure of the extent to which and the reasons why a person feels as if she is in a virtual world. Social presence, according to Heeter, refers to the extent to which other beings both living and synthetic exist in the virtual world and appear to react to the user. Finally, environmental presence refers to the extent to which the environment itself appears to know that you are there and react to the user.

3.4 Comparing the concepts awareness and social presence

In the area of computer supported cooperative work (CSCW) people's level of awareness of others activities has been identified as important for cooperation in groups. The concept of awareness is generally used in terms of, individuals' perception of others' activities and the status of others' work-processes. When people that cooperate do not have the opportunity to get this kind of information, if they for example work in a distributed way, studies have shown that they do not reach the same quality in joint projects (Kraut et al., 1993). Awareness is necessary for all kind of cooperation but it can vary due to what degree of focus the cooperation has (Gaver et al., 1991). When working closely together in order to solve a task, individuals get a large amount of information about status in the work process and thereby get a high degree of awareness. A lower degree of awareness characterises cooperation where a main task is divided into subtasks between individuals. An even lower degree of awareness characterises occasional encounters where no task is to be accomplished. Apart from this, co-located individuals have general awareness of events going on in the surrounding context. In the case of distributed collaboration, this general awareness has to be transmitted indirectly through the media, and thus differs depending on the media used.

The concept awareness is sometimes confused with the concept of social presence. As explained earlier, social presence refers to the feeling of being present with another person at a remote location (Short et al., 1976). Then what is the difference between social presence and awareness? Well, the most important difference is that social presence is a subjective perception of a system's capacity to mediate social information and the perception of the social quality of the interaction between two persons. The concept awareness on the other hand concerns people's knowledge of other people's activities that facilitate the coordination of their own activities. Social awareness is the awareness of the fact that another person is present, not the perception of the extent of social information that one gets from that person regarding that persons emotive state as is the case with social presence.

Another aspect that presumably differs for the two concepts, is whether one must get continuous information or not in order to maintain awareness or social presence respectively. It is not likely that people maintain awareness of others actions, if all information of the other persons' activities are removed. Short et al. (1976) however, argue that the feeling of social presence is maintained, if one for example shuts ones eyes and plugs ones ears when in a video conference, because one is still affected by the knowledge that the other person is seeing and hearing you. This explanation is similar to the ideas about

social facilitation that the conviction of someone else's presence is enough for a person's behaviour to be changed.

If high awareness settings are compared to settings where people feel high social presence it becomes clear that situations where people have good awareness are not always situations where people feel social presence to a large extent. One example is a shared graphical user interface (GUI) where people see the other person's proxy and in that way the other person's actions. In that situation the person has good awareness of the other's actions but at least hypothetically a low feeling of social presence. It is however hard to think of a rich medium that does not convey good awareness of others' activities. Possibly, a telephone conference might serve as an example. In a telephone conference people do not notice very well if a person engages in multitasking while managing the conversation. However, rich media that support different human senses like vision, hearing, touch usually add to people's awareness of each other as well as to the social presence.

3.5 The contribution of each perspective

The aim of this chapter has been to present a number of theories that make statements, about to what extent humans perceive that media differ and the effects that this might have on mediated interaction. A general statement among all theories is that people perceive that media differ regarding media richness, social presence or presence. Another general argument is that the level of media richness, social presence or presence increases when more modalities are provided. The major difference between the theories discussed in this paper is that structural contingency theory, media richness theory and social presence theory are mostly focused on perceived qualitative differences in mediated interaction between people. The research area in virtual reality however, that investigates presence, is more focused on perceived qualitative differences in the human-computer interaction. Recently however, concepts such as "togetherness" or "co-presence" have appeared that address social aspects of virtual environments. The operationalisations of those concepts show that they come from the presence tradition as they focus more on the perceived realness of other persons, an aspect that is never questioned in the social presence approach.

Structural contingency theory, media richness theory and social presence theory all claim that people are aware of differences in media and that they actively choose the medium they perceive as appropriate in a certain communication situation. The presence research in the area of virtual reality does not make this specific statement. In addition structural contingency theory and media richness theory make predictions on an organisational level, suggesting that

more complex and undefined work needs richer communication channels. In contrast, the focus in social presence theory is on a task level and not on an organisational level. Social presence theory predicts that the outcome of tasks involving attitude change or negotiation or formation of impressions of strangers is improved if the communication channel provides a high sense of social presence.

Media are graded along a continuum in both media richness theory and social presence theory in the order from higher to lower in capacity. Short et al. (1976) did not regard the order of media he presented as conclusive, as it can depend on conditions of testing and context. The order that is suggested regarding the capacity to transmit social presence is from higher to lower, face-to-face, television (video), multispeaker audio, telephone (also speakerphone and monaural audio) and finally business letter. Daft (1998) finds that the order of media regarding media richness is from higher to lower, face-to-face, telephone and voice-mail, written addressed documents like letters and faxes and written impersonally addressed documents. This kind of ordering of media can be problematic, as new combinations of media and support for more human modalities can overthrow it, and the media richness theory has been criticised (El-Shinnawy and Markus, 1997). Short et al. (1976) suggest that there is a difference between, to what extent people think that qualities and characteristics of media affect the efficiency of communication and collaboration, and the extent to what performance in task accomplishment actually is affected. The subjective experience of a system's appropriateness, social and interactive properties, still has important implications for people's desire to use different systems, and therefore has to be considered carefully.

4 Mediated communication

Communication through media has not always been multimodal in the sense that the technology provides information to many senses simultaneously. Traditional media like letters are only visual and telephone is only auditory, whereas videoconference provides both an image of the other person's mimics and gestures as well as the sound of the voice. Other forms of technology like virtual reality can also make it possible to move around in an environment and to position an embodiment in the direction where one's focus of attention is. Haptic interfaces can provide tactual feedback that makes it possible to touch and modify graphical objects as if they were physical. Even smell can be provided, in for example an interface that gives off a scent that can be sensed by a person in a room at one geographical location when a family member picks up an object at another location (Strong and Gaver, 1996). The effects of supporting each modality in mediated human-human interaction are not fully understood, and even less is known about the effects of providing combinations of modalities. Theories of social presence (Short et al., 1976; Rice, 1993) have argued that the more modalities used for mediated communication the more socially rich the interaction is perceived to be. Research on virtual environments has also argued that the more modalities used, the more present and immersed a person perceives herself to be in a remote or distributed environment when she in fact is physically in another environment (Witmer and Singer, 1998).

4.1 The impact of communication medium on collaboration

A large number of studies have investigated benefits and drawbacks with communication via different media such as email, telephone, video conference and even written letters, sometimes using face-to face as a baseline to compare with. The intuitive hypothesis is often that richer media would improve both performance and satisfaction when people collaborate remotely. However, this assumption is not as straightforward as hypothesised, as many studies have shown. It has been largely established that voice communication makes the most significant difference compared to text communication across a large variety of contexts regarding how efficiently synchronous collaboration can be performed and the group's ability to form social relations (Ochsman and Chapanis, 1974; Chapanis, 1975; Vaske and Grantham, 1989; Short et al., 1976; Sellen, 1992; Matsuura et al., 1993). Not so large benefits have been found when video is added compared to communicating with voice only, even though social psychological communication theories put great emphasis on the hypothesised importance of non-verbal cues for our understanding of other persons' intentions and emotions (Short et al., 1976; Rutter and Stephenson, 1977; Clark and Brennan, 1991).

4.2 Communication by text-chat

It is frequently found that groups that communicated via text produce less amount of words in total and that the communication has higher focus on the task rather than on social relation building (Ochsman and Chapanis, 1974; Chapanis, 1975; Short et al., 1976; Straus, 1997). Clark and Brennan (1991) argue that because text lacks "co-presence", visibility and audibility of the communication partner and because the information sent is not received as it is produced and that the communication does not necessarily follow a turn-taking sequence, it takes more resources to achieve common ground. The principle of least collaborative effort says that "in conversation, the participants try to minimize their collaborative effort – the work that both do from the initiation of each contribution to its mutual acceptance" (Clark and Brennan, 1991). Clark and Brennan argue that text is more costly to produce and that more effort is put into the formulation of a message as it is reviewed by the receiver and might be saved. It takes longer time to receive an answer in text which slows down the collaboration process. It is often harder to develop a common understanding of the context of a text message as it is difficult to convey nuances in information without tone of voice, mimics or gestures. It is also hard to repair mistakes in text communication. Text communication does however have the benefit of being reviewable as it is easy to read something

many times. A text message can also be revised many times before the message is sent which makes it easier for the sender to formulate and for the receiver to understand complex information.

A result of not getting rich social cues when for example communicating with text-chat can be that group members become less socially connected which induces a sense of anonymity and feelings of depersonalisation (Kiesler et al., 1984; Kiesler et al., 1985). Depersonalisation can be defined as a feeling that influences our perception of others, making them seem "more nonpersons, more machine-component like, more object-like, and so on" (McGrath, 1984). In a communication situation where group members feel depersonalised they are less motivated to share personal information and to inquire about others (Kiesler et al., 1985). Less supportive communication, like praise or encouragement, was found when interaction in lean and rich media (telephone and FTF respectively) was compared (Stephenson et al., 1976). Another sign of depersonalisation is that communication is less inhibited in groups that communicate with text compared to richer media. When no social context cues help members to monitor discussions and express emotions in a rich way, group members tend to criticise and even insult each other more often, which is also called "flaming" (Sproull and Kiesler, 1986).

However, Straus (1997) recognizes that although the amount of uninhibited communication has been found to be significantly higher in text communication situations than in face-to-face communication, the actual number of for example "flamings" is usually low. Straus argues that it is necessary to separate between task-relevant disagreement and personal attacks. Rather than arising from antipathy, expressing explicit disagreements and superlatives might be needed to a larger extent in text communication in order to compensate for the loss of emphasis provided by nonverbal cues. In a study she found higher rates of disagreement for groups that communicated via text than face-to-face groups but no differences were found in the incidence of personal attacks between conditions. Interestingly, it was found in this study that groups that communicated with text produced higher rates of supportive communication. This finding casts doubt on the ideas about depersonalisation and instead supports Straus (1997) idea that superlatives are probably used to compensate for the loss of emphasis otherwise provided by nonverbal cues.

The effects of communication mode on group cohesiveness have also been studied. Group cohesiveness is multidimensional in nature consisting of interpersonal attraction, attraction to the group task and group pride. Straus (1997) examined interpersonal attraction in her study and found that groups that communicated with text reported lower cohesiveness than face-to-face groups.

Text communication increases anonymity and reduces inhibition and this has been found to result in a more equal participation rate in text communication (Kiesler et al., 1985). When individual differences that are associated with social status cannot be perceived, group members that otherwise feel inhibited might communicate more. Straus (1997) found in her study that participation rate was more equal among group members who communicated with text than in face-to-face communication. Social inhibition also occurs in shared virtual environments. It was found that subjects performed a novel task more poorly with the presence of another avatar, but only if the avatar was recognised as embodied by a human (Blascovich, 2002).

4.3 The added value of voice communication

Among the most robust findings is that high quality audio makes a significant difference for mediated synchronous collaboration (Ochsman and Chapanis, 1974; Chapanis, 1975; Short et al., 1976; Fish et al., 1990; Tang and Isaacs, 1993; Jensen et al., 2000). It is generally found that providing a audio channel to collaborating group members results in significantly faster task performance compared to a text communication condition (Chapanis, 1975), that more words are spoken and that better quality in the solution is achieved. Audio is generally perceived to be a significantly richer communication medium than text, and voice communication is therefore preferred over text for performing complex tasks involving ambiguity, conflicting interests and relation management (Short et al., 1976; Kraut et al., 1992; Rice, 1993; Daft, 1998).

In a study where ten different communication modes were compared, the most significant conclusion was that including voice communication makes a very large difference (Ochsman and Chapanis, 1974). The communication media compared were: typewriting only, handwriting only, handwriting and typewriting, typewriting and video, handwriting and video, voice only, voice and typewriting, voice and handwriting, voice and video and finally a communication rich mode that included all of the channels above. The tasks were chosen to be credible "real world" tasks and consisted of a class scheduling problem, a fault finding problem and an object identification problem. The most striking result was a clear dichotomy between the time spent to complete the tasks in the voice modes compared to the text modes. The voice modes were significantly faster than the text modes. Furthermore, the video and voice mode was slightly faster than the voice only mode and the typewriting only mode was almost identical to the video typewriting mode. The number of messages communicated was also analysed and a similar dichotomy between the voice modes and the text modes was found as for task performance time. Significantly more messages were communicated in the voice modes alhough

tasks were performed significantly faster in the voice modes than in the text modes. No significant difference was found for adding video.

The conclusions are, that the result that voice modes are more efficient than text modes is robust as it holds for all three kinds of tasks. The next conclusion is that the addition of a video channel has no significant effect on communication times or communication behaviour. Finally it was concluded that there are no differences in efficiency between handwriting and typewriting. The large difference between text and the voice condition in task completion time and amount of words used, is probably explained to a large extent by the costs of writing text compared to speaking (Clark and Brennan, 1991).

In other studies (Chapanis, 1975), various modes of interaction were examined ranging over voice, handwriting, typewriting, video/voice and face-to-face communication. Similar results compared to Ochsman and Chapanis (1974) study were found, that people produced more words in the voice condition than they wrote in the text condition, and people completed tasks faster when voice was provided compared to when text was used. Therefore the rate of communication, measured by dividing number of words with the time spent, is much higher when people speak than when they write. It was found in four different experiments that people use more words in total in face-to-face communication than in voice only communication, but the difference was not significant (Chapanis, 1975).

It has also been found that voice communication is important for people's willingness to cooperate and feelings of trust in online environments. Jensen et al. (2000) examined the nature of cooperation between two players of an electronic continuous version of the Prisoner's dilemma game under four different conditions: no communication channel, text-chat, text-to synthesized speech, and voice. They found that immediate forms of communication e.g., voice are important for cooperation and trust in online environments. It is very tempting in the game that was played by the subjects to behave selfishly but the best overall outcome for both persons is obtained if both players cooperate. The best possible outcome for an individual is however, not to cooperate at all while the other person cooperates. The result showed that the pairs of subjects that communicated via a voice channel showed significantly greater levels of cooperation than all other conditions. Furthermore, players rated the other player's intelligence significantly higher in the condition where the cooperators spoke to each other than in all the other conditions. The results showed that players also cooperated slightly more in the text-to speech condition but the result was not significant.

4.4 Does video make any difference for remote collaboration?

Research has not found as uniform results from using video conferencing, as for text and voice communication, regarding benefits or disadvantages. On the contrary, several studies have shown that adding video does not make a significant difference compared to voice communication on task performance measures (Egido, 1990; O'Connaill et al., 1993; Anderson et al., 1996; Olson et al., 1995; Finn et al., 1997; Whittaker and O'Conaill, 1997). Little value was found when adding a visual channel for task-based communication such as information transmission, or collaborative problem solving (Chapanis 1975; Williams, 1977). In his studies, Chapanis (1981) found that in terms of time to complete tasks and amount of words spoken, there was no difference when a visual channel was compared to a voice only condition. It seems that the only tasks for which video has been shown to add value are certain kinds of negotiation tasks (Short et al., 1976). However, a substantial number of studies report that subjective ratings of satisfaction, social presence, affiliation and other dimensions do differentiate between especially voice and video/voice communication channels (Short et al., 1976; Olson et al., 1995; Daly-Jones et al., 1998). One study showed that a person's preference ratings for how much he liked other persons whom he met in an voice mode were lower than the ratings for those whom he met in a video mode or face-to-face (Williams, 1977). Video has also been shown to support informal communication and relation building (Dourish and Bly, 1992; Bly et al., 1993; Kraut et al., 1993).

Little benefit of video conference on task performance quality was found compared to using only voice when groups of three people were performing a system design task (Olson et al., 1995). The subjects in this study knew each other from before because they had worked together in their everyday work context. During the experiments they were seated in three different rooms. The results showed that the judged quality of the solutions of the design task did not differ whether the groups used voice only communication or if a video channel was added. No differences were found between the face-to-face, video/voice and voice only conditions, regarding the total time or amount of words produced when solving the task and using a shared editor. However, the video/voice groups spent significantly less time stating and clarifying the issues than the voice only groups. When ratings of satisfaction were analysed it was found that voice only groups were significantly less satisfied with the quality of the discussion than video/voice groups. Remote groups with only voice communication reported that they were less able to tell how their other group members were reacting to things said than the video/voice groups, and they reported that the communication system got in the way of their being able to persuade others about their ideas or to resolve disagreements.

In one study it was hypothesised that non-native speakers would have more use of a video channel than native speakers when solving a task. The task was to find the way on a map together with another person (Veinott et al., 1999). The results showed that the performance in time and precision did not differ for the native speakers between a video/voice and a voice only condition whereas non-native speakers did benefit significantly from video. Performance in the video condition was equally good for both native and non-native speakers but non-native speakers performed worse in the voice only condition.

In a study by Anderson et al. (1996) it was found that there was no advantage of adding a video channel to voice communication when subjects performed a task involving planning a trip using a map with the help of a human travel agent. Task performance was equally good in a voice only condition, a video/voice condition and in a face-to-face setting. It was found that dialogues were significantly longer in the voice condition than in the face-to-face condition. There was no significant difference between a voice only and a video/voice condition in the length of dialogues. A similar pattern was found for the number of changes that was needed in order to make the solution more optimal. Again, there were significantly more optimal changes of plans in the face-to-face than in the voice only condition but adding video again failed to deliver the same benefits to users as the face-to-face interaction. When subjects' ratings on the question, "how aware they were of the travel agent" were analysed the video/voice communication was rated significantly higher than the voice only setting. Anderson et al. (1996) argue that this question measured social presence as well as another question that gave the same result "how often they worried that they had lost contact with the travel agent". I argue however, that it is uncertain if this type of questions is on par with social presence theory. It is nevertheless an interesting result that subject's ratings once again prove to differentiate between voice only and video/voice whereas behavioural data such as task performance do not.

The prediction that conversations would be more fluent if video was provided compared to discussions conducted over a voice only link was not verified in a study by Daly-Jones et al. (1998). No differences were found in the number of conversational turns, the length of spoken turns or the amount of overlapping speech. It was hypothesised that people would ask more questions in the voice only than in the video/voice condition because of two reasons. Firstly, questions represent a way of securing mutual understanding and agreement. Secondly, they provide a way to coordinate speaker transition. The results from the study confirmed that significantly more questions were asked in the voice only condition. Finally, Daly-Jones et al. (1998) found that users rated on a questionnaire that they were significantly more aware of the presence of their partner in the video condition, could monitor their partner's attentional status better, and felt that video communication aided collaboration. Subjects felt that

it required significantly more effort to converse in the voice only than in the video/voice condition.

Short et al. (1976) found that communications media could be ordered in a continuum based on subjects' ratings of social presence on bipolar scales. Subjects perceived that social presence was highest in face-to-face communication and that communication mediated by television (video), multispeaker audio, telephone (also speakerphone and monaural audio), and finally business letters mediated lower levels of social presence in the order mentioned. Short et al. concluded that that visual cues generally increase perceived social presence and that there is generally a significant difference between video/voice communication and voice only communication. Another important result was that adding voice communication generally increases perceived social presence significantly compared to written messages. Short et al. (1976) report mixed results regarding differences in social presence between video/voice and face-to-face communication and suggest that these depend to a large extent on the experimental conditions like for example the size of the video picture. Meetings involving the task of managing conflict between the participants are also more sensitive to the extent that the medium determines feelings of social presence. Negotiations, bargaining, persuasion in order to change attitudes, coalition formation or getting to know strangers are activities in which more conflict is potentially present. The outcome of these types of meetings are more dependent on the participants' salience of each other, feelings of social presence, and in such meetings video/voice has proven to be more efficient than voice only for task performance. Face-to-face meetings however, are perceived to be the optimal for such meetings. It is on the other hand generally found that activities such as information transmission and cooperative problem solving are activities in which interpersonal relationships are relatively unimportant. In these cases no advantages are usually found when video is added.

4.5 Human-human interaction in virtual reality environments

In collaborative virtual environments (CVEs), the communication mode is often text-chat, whereas audio or video channels are less often used. For this reason, it is interesting to compare the use of different kinds of communication media in CVE´s.

Becker and Mark (2002) studied three different online environments, Active Worlds in which people communicate through text and have a graphical embodiment called an avatar that looks like a human, Onlive Traveller in

which people can speak to each other and use an avatar looking like a head, and LambdaMOO in which all representations of users and communication are text based like in an interactive book. Becker and Mark, apply social presence theory in order to explain the finding that social conventions differ in the three virtual environments. They hypothesise that social conventions are more binding in virtual environments that induce stronger perceptions of non-mediation and thereby stronger feelings of social presence. People use visual representations in the form of avatars in both Onlive Traveller and in Active Worlds that make it possible to use the position of the avatar to for example convey focus, group formation and movement in space. Voice communication that has been shown to increase social presence is only provided in Onlive Traveller. In Onlive Traveller avatars have standard expressions such as eye-blinks and lip movements and in Active Worlds avatars have standard gestures such as jumping out of joy, fighting and so on. In the LambdaMOO environment, communication and representations of users are text-based. The lack of visual avatars or voice communication makes emote commands necessary in order to express emotions.

The empirical observations showed that greetings are important in all three environments, but that the way they were made differed both because of interface design and the available communication media. Greetings were performed in a more elaborate way in environments with voice communication than when only text communication was provided. In the voice environment people used their avatars to navigate to a specific person or group of people and positioned their avatars so that they "faced" the other or waited to be invited in the circle that groups often formed when talking. This replicates the way people behave in a real world face-to-face situation. This convention that people position avatars face-to-face during interaction was also found by Bowers et al. (1996) in a virtual environment where people used voice communication. In both environments where text communication was used, people made public greetings to everyone present when they entered the virtual worlds (Becker and Mark, 2002). In the Active Worlds, in which people have a graphical embodiment, people moved close to a person that they communicated with only about 30% of the time. Social presence theory predicts that voice communication increases the feeling of social presence. Becker and Mark (2002) argue that this explains the fact that users put a lot of energy into positioning their avatar in a way that would have been efficient in a real world setting. However a counterargument could be that this is instead due to the characteristics of the audio that was stronger if a person faced the other person and also that voice communication might be more disturbing to others than text if it is sent in parallel to others' utterances.

Commitment to a speaking partner differed in the three environments. The social obligation to remain for some time was stronger when people

talked to each other and possibly also because of the added cost involved in positioning the avatar face-to-face. When text was used for communication speaking partners appeared to change more often. Becker and Mark, (2002) suggest that the reason for this was that it was easy to address a new person simply by typing in a new avatar name in the public text window. In the voice setting people generally exchanged a farewell when they parted but in the text environments this was not so common, farewells were said to the whole group in the text setting when someone left. Becker and Mark, (2002) argue that voice communication probably increased the feeling of social presence in the Onlive Traveller environment and that the behaviour that people faced each other explained why they became more committed to their speaking partner than they did in the Active Worlds. It was also found that people repositioned their avatars in e.g. a circle to indicate that they were a group that communicated when they used voice communication, whereas group formations could only be worked out by looking at the text log when text was used for communication. In the two environments that had embodiments, private conversations were achieved by moving away to a less populated area or by positioning avatars closer or even up side down. In the environment that did not use embodiments a whispering mode was used for private conversations. It seemed as if people did not identify as much with their avatars in the Active Worlds as in Onlive Traveller, maybe because they did not rely as much on the avatars for the conversation that was in some ways separated from the avatars. One example of this was that avatars were not repositioned by people in Active Worlds when conversation partners changed. Becker and Mark (2002) argue that this is a sign of lower feelings of social presence. It is hard however to justify such a strong conclusion, the reason for this difference in behaviour might just be that it takes quite some effort to move the avatar around, and that it was not necessary for efficient text chat communication in Active Worlds.

Signs of discomfort when avatars came too close to another avatar in both Onlive Traveller and in Active Worlds indicated that interpersonal distance was important in these environments just as it is in the real world. People verbally asked others to move away or reposition themselves if social distance was perceived to be violated, and people responded by moving away, or turning to the side which is what often happens in elevators in the real world when space is limited. Slater and Steed (2002) found in their study that people are very mindful of the avatars of others, avoid passing through other avatars and try not to carry out actions that would cause distress in others. Becker and Mark (2002) do not seem to find a difference between Onlive Traveller and Active Worlds on this point and therefore argue that level of social presence does not affect this aspect. Their explanation for this is that both environments had avatars and that they were equal on that point. However, if a person identifies more with their avatar, because of the spatialised audio system, as in Onlive

Traveller it could be argued that they should have been more sensitive to violation of their avatar's space according to social presence theory. Balscovich (2002) for example showed that people moved closer to an embodiment that had its eyes closed (a sign that the person could have left the session temporarily) than an avatar that was looking and blinking. If people did not utilise their avatars as much in Active Worlds and the avatars thus in a sense were less inhabited, which was shown by the fact that it was not moving as much, then social presence could have been lower and thus annoyed responses might be less frequent. It is very plausible that even if people responded to violations in both environments one would, if analysing this in more detail, find that violations of personal space have different implications depending on the level of social presence. The direction of causality is not evident however, low social presence might result in people being more frightened by someone coming close than if people perceive high social presence in the interaction. Also, one probably has to distinguish between how embodied an avatar is perceived to be, something that might vary during interaction, and the level of social presence that the particular medium is perceived to mediate that is probably more stable.

Slater and Steed, (2002) found that if a person is instructed to consistently exhibit a specific behaviour like monitoring another, it is noticed by others and is interpreted in many different ways like a bug in the system, as intimacy or destructive behaviour. They also found that how avatars look has an impact on relationships, especially if there is a difference in the level of realism. A realistic avatar is perceived as a more important person. A person having a more realistic avatar hesitates to have an extensive dialog with cruder avatars until they show human behaviour like laughing (Slater and Steed, 2002). Finally if one person has more sophisticated technology, like if being immersed using a VR-helmet instead of using a desktop, that person is significantly more often perceived as the leader by others (Slater et al., 2000; Slater and Steed, 2002). The interpretation of this result could be that the person using more sophisticated technology is more efficient in interacting in the environment and thus seems to be more able to lead than the others. Another explanation could be that the person using more sophisticated technology feels present to a larger extent than the others and becomes more engaged in the task, and takes command to a larger extent, which the others recognise.

A slightly different approach was taken in this thesis compared to the studies mentioned above (Sallnäs, conditionally accepted). An experiment was conducted comparing the effects of using different communication media when interacting in a virtual exhibition that was developed using Active Worlds. The communication media compared was: text chat that is a feature in Active Worlds, a telephone connection with head sets, and a video conference using 21 inch television monitors as well as a telephone connection with head sets. In

this study social presence was measured by a questionnaire that was developed by the author but the questionnaire items are inspired by the Social presence theory (Appendix 2 in article C) and they are to some extent equivalent to items used by Short et al. (1976). The definition of social presence in the study was "feeling that one is present with another person in a shared environment". In this experimental study presence defined as, "feeling as if being in a mediated environment" was referred to as virtual presence. Virtual presence was measured using Witmer and Singer's (1998) Presence questionnaire (PQ). The results showed that people rated their perceived social presence and virtual presence to be significantly lower if they had communicated with text chat, when they solved the task together in the Active Worlds exhibition, compared to the other two conditions. No significant difference for these measures were found whether video was used in addition to voice communication or not in the Active Worlds exhibition. This result supports the argument that adding voice communication in a virtual environment like this makes a difference. These results were compared to results from a follow up study were people performed the same task together, using a two dimensional web environment instead with either voice or video/voice communication. It was found that adding video in the web conditions significantly increased virtual presence but this was not the case in the Active Worlds condition. The same pattern was found for social presence but the result was not significant. This suggests that a shared virtual environment such as the Active Worlds is either so involving or distracting that the video is not beneficial or that the virtual environment to a certain extent has equivalent features as the video medium. Such features could for example be those that other researchers have pointed out, that people can face other avatars in virtual environments, that focus and level of attention can be indicated by the position and movement of the avatar, and that interpersonal distances can be managed with the avatars (Balscovich, 2002; Mark and Becker, 2002; Slater and Steed, 2002).

It was also found that it took significantly longer time to complete the task in the Active Worlds exhibition when text chat was used compared to the other conditions and that significantly fewer words in total was used in text chat (Sallnäs, conditionally accepted). No differences were found however, between the voice and the video/voice conditions regarding the time spent or the amount of words used. But, when words used per time unit was analysed, it was found that the tempo in the dialogue was significantly lower in the video/voice condition than in the voice only condition, and that the tempo in the text chat dialogue was significantly lower than in both the voice and video/voice conditions. When these results were compared to the results from the follow up study, it was found that the tempo in the dialogue was faster in the web voice and web video/voice conditions than in the CVE voice and CVE video/voice conditions in Active Worlds. Finally, it was found that people completed their

task significantly faster in the web voice and CVE voice conditions than in the web video/voice and CVE video/voice conditions.

These results are not easy to interpret but they show that text chat is different compared to using telephone or a video conference for the measures used in this study. Furthermore, the results show that it makes a difference for the tempo in the dialogue if video is added to voice communication in a virtual environment. No other advantages were found of adding video in a CVE. However, when a comparison was made with interaction in a web setting the picture became more complicated. It can at least be concluded, that both communication medium and whether a web or CVE is used makes a difference for people's experiences and behaviour in these kinds of settings.

5 Haptic force feedback in mediated interaction

In face-to-face communication and collaboration people are used to being able to both see and hear other persons. People also take for granted the possibility to give objects to each other, to shake hands or to get someone's attention by a pat on the shoulder. However, most systems for mediated collaboration do not take physicality into account. Now emerging media space technologies like three-dimensional haptic interfaces makes it possible to interact physically in shared haptic object spaces. Many questions then arise about the effects of these modalities on communication and collaboration.

5.1 The sense of touch

The perception of touch is complicated in nature. The human touch system consists of various skin receptors, receptors connected to muscles and tendons, nerve fibres that transmit the touch signals to the touch centre of the brain, as well as the control system for moving the body. Different receptors are sensitive to different types of stimuli. Tactile perception is defined as perception mediated solely by variations in cutaneous stimulation (Loomis and Lederman, 1986). There are receptors sensitive to pressure, stretch of skin, location, vibration, temperature and pain. Contrary to what one might think, there does not seem to be one receptor type for sensing pressure, another for sensing vibration and so forth. Rather, the different receptors react to more than one stimulus type (Burdea, 1996). The skin on different parts of the body is differentially sensitive

to touch. The ability to localise stimulation on the skin depends on the density of the receptors, which are especially dense in the hands and face. Moreover, a great deal of information provided by the kinesthetic system is used for force and motor control. The kinesthetic system enables force control and the control of body postures and motion. The kinesthetic system is closely linked with the proprioceptic system, which gives us the ability to sense the position of our body and limbs. Kinesthetic perception is defined as perception from joints and muscles, by limb movement alone, of hardness, viscosity and shape (Loomis and Lederman, 1986). Receptors (Ruffini and Pacinian corpuscles, and free nerve endings) connected to muscles and tendons provide the positional information. Haptic sensing is defined as the use of motor behaviours in combination with touch to identify objects (Appelle, 1991). In haptic perception both the cutaneous sense and kinesthesis convey significant information about distal objects and events. The haptic system unifies input from many sources, e.g., position of fingers, pressure, into a unitary experience.

Manipulation of objects can take many forms and one taxonomy illustrates how diverse functions haptics fulfils in everyday life (Lederman, and Klatzky, 1987). People use different strategies depending on the purpose of the tactile manipulation, such as investigating the weight, form, texture or softness of an object. Joint manipulation of objects can take just as many forms. One example is jointly grasping an object and moving it through an area that might have restrictions (Ruddle et al., 2002). Another example is moving an object by pushing from both sides and lifting the object together. Yet another type of joint manipulation is grasping an object and handing it to another person.

In a shared haptic object-space people can coordinate joint movement of objects by signalling direction through haptic force and they can give objects to each other almost without verbal communication. Also, the bilateral (Biggs and Srinivasan, 2002) qualities of haptic perception make it possible to both move an object and to get information from it at the same time. In a collaborative situation these aspects of haptic sensing might facilitate the joint understanding of complex information or how something is constructed.

5.2 Psychology of touch

The use of the sense of touch for understanding information in the form of texture and shape is often neglected in computer interface design because of the traditionally perceived dominance of vision for interacting with graphical objects. Vision is usually regarded as the dominant sense for judging distance. Berkeley however, argued that "the ideas of space, outness, and things placed at a distance are not, strictly speaking, the object of sight…I neither see distance itself, nor anything I take to be at a distance" (Heller and Schiff, 1991).

Berkeley suggested that we learn to associate visual and auditory experiences with tangible ideas. In this way one can claim that touch to a certain degree educates vision and audition. Researchers do not agree on this topic, recent data indicate that in fact vision educates touch, not vice versa (Katz, 1989). Distance might nevertheless be an ungraspable and abstract phenomenon, if not experienced through body movement and touch.

Touch has by a number of philosophers been seen as dominant over other senses in terms of an existence proof for objects, that is, we test reality of a mirage or illusion by trying to touch it (Heller and Schiff, 1991). Humans tend to think of touch as the "reality sense" because we know that it is relatively easy to fool vision by distorting lenses, differences in lightning and viewing conditions. Traditionally touch has been dismissed as a lower sense whereas vision and hearing are looked upon as the higher senses. Katz (1989) however, argued that touch from a perceptual viewpoint must be given precedence over all other senses because its perceptions have the most compelling character of reality. Katz argued that:

"touch plays a far greater role than do the other senses in the development of belief in the reality of the external world. What has been touched is the true reality that leads to perception; no reality pertains to the mirrored image, the mirage that applies itself to the eye."

Other senses are more ambiguous than touch and therefore touch is often used to check on reality, like when a rod is held immersed in water appears broken to the eye the hand makes sure that it is not. It is hard to imagine that we would believe what we see rather than what we feel. Most people think that an object is rather stable over time regarding its size and shape. This is probably a consequence of the fact that, even though the retinal size and shape of an object can differ due to viewing conditions, angles and distance, the touch percept is more or less stable. We think that an object has only one true size and shape and only one true surface structure. The fact that people generally perceive touch percepts to be stable becomes a very important aspect to consider when designing haptic interfaces. Because people trust the haptic perception, and are usually not used to simulated haptics, inconsistencies in the haptic simulation can have serious consequences. One problem is that people explore what they see to a larger extent than things that are invisible but haptically perceivable. This means that, if great care is taken to design a complete haptic object, but the visual graphics for example only reveal parts of the haptic model of the object, the risk is large that only the visible parts will be explored. Some haptic illusions can also surface because perceptual events that are very infrequent in the real world can be easily simulated. One example is that if two boxes with different sizes but equal weight, that seem to be of the same material, are lifted by a person, the larger object is perceived to be lighter than the smaller one.

This is because in nature a larger object should be heavier than a smaller one if they are of the same material.

For the very influential perception psychologist Gibson (1979), the man-made aspects of the environment like for example tools should not be seen as separate artificial entities but as a part of the natural environment, only that it is modified by humans. Tools still consist of the same fundamental material as everything else in our surroundings. Gibson (1979) also argues that all aspects of the world provide affordances for humans as well as animals. Ground affords support for walking, air affords breathing, water affords drinking and solid materials afford manipulation by the human body and primarily the hands. Depending on the qualities that a solid material has it affords different kinds of manipulation and different things can be manufactured, usually fabricated by hand. Gibson (1979) argued that:

"To identify the substance in such cases is to perceive what can be done with it, what it is good for, its utility; and the hands are involved"

Gibson (1979) gives a number of examples of affordances that different objects have that depend on their properties or qualities: colour, texture, composition, size, shape and features of shape, mass, elasticity, rigidity and mobility. An elongated object of moderate size affords wielding, hitting, or raking. A graspable rigid object affords throwing and an elongated elastic object affords binding or weaving. In contrast to many other psychologists Gibson thought that phenomenal (psychologically perceived) objects are not built up of easily discriminative parts or qualities but are instead perceived as integrated unified entities that afford certain behaviours. We identify an object as one whole entity, one specific thing, not as a bunch of separate qualities. Gibson (1979) argued that:

"The meaning of an object is perceived before the substance, surface, colour and form, are seen as such. An affordance is an invariant combination of variables, and one might guess that it is easier to perceive such an invariant unit than it is to perceive all the variables separately. It is never necessary to distinguish all the features of an object and, in fact it would be impossible to do so. Perception is economical. Those features of a thing are noticed which distinguish it from other things that it is not - but not all the features that distinguish it from everything that it is not"

Among researchers that study the tactile sense, the importance of movement in relation to touch perception has been recognised (Gibson, 1979; Katz, 1989). Gibson thought that movement was essential for perception, the movement of the limbs and head relative to the body and the locomotion relative to the environment. Accordingly, Gibson makes a distinction between passive and active touch. Touch is passive when the person does not move and information

is imposed on the skin (Heller and Schiff, 1991). Active touch consists of self-produced movement that allows the perceiver to obtain objective information about the world. It was shown that people rely a lot on explorative movement to recognise shapes when blindfolded. In an experiment Gibson (1979) found that when "cookie cutter" shapes were pressed into the palm of the hand of the subject (passive touch) the shape recognition was as low as 29% whereas recognition was 95% when subjects could explore (active touch) the shape freely (Appelle, 1991). Visually impaired people however, in many cases rely on passive touch for reading by using a vibrotactile display where the index finger is passive and information is projected on the fingertip. In essence it is generally argued that haptic perception is active touch as information is obtained through both tactile perception by the nerves in the skin and kinestetic perception by nerves in the muscles and joints. In the use of haptic interfaces, touch is usually active rather than passive.

Gibson (1979) argued that humans not only perceive the affordances of objects but that also the social behaviour of other beings, including animals, have affordances. Humans are dynamic and convey complex patterns of behaviour that other humans interpret as affording certain behaviours reciprocally in a continuous fashion. Humans interact with one another and behaviour affords behaviour according to Gibson (1979). Nurturing behaviour, fighting behaviour, cooperative behaviour, economic behaviour, political behaviour – all depend on the perceiving of what another person affords, or sometimes the misperceiving of it. Gibson (1979) argued that:

"The perceiving of these mutual affordances is enormously complex, but nonetheless lawful, and it is based on the pickup of the information in touch, sound, odour, taste and ambient light"

Following this line of reasoning, it is evident that the multimodal input of information is important for an accurate understanding of another person's social affordances. In mediated interaction only a selection of a person's affordances in the real world can be conveyed to a receiver. In for example text-only communication a person can only communicate the message in text and the receiver of the message has to imagine for example emotions through the text description. Another person's voice conveys much more detailed social affordances through pitch, loudness, tempo and melody. Video conveys even more communication behaviour that hypothetically would improve social affordances but video is sometimes more unreliable than audio as the optic system can distort many of the social signals such as eye gaze, body size and distance.

Haptic feedback is taken for granted in our non-mediated interaction with others. Tactile contact with others is managed in very subtle ways because of

the fundamental importance of protecting ourselves from harm at the same time as tactile contact is probably essential for well-being and survival. Gibson (1979) includes the importance of all senses for perceiving social affordances of others as well as for perceiving affordances of objects around us. Mehrabian (1972) includes touching as the most important variable in his construct "immediacy" along with interpersonal distance, forward leaning toward the addressee, eye contact and body orientation in that order of importance. Immediacy is one of the concepts that influenced Short et al. (1976) in their work on the social presence theory. However, haptic feedback is just starting to be used in interface design for interaction with graphical objects and has been used even less for mediated human interaction and joint manipulation of objects. Tactile contact in real encounters provides the most fundamental proof of something being real and believable. This is probably also true in social interaction even when the haptic interaction is very limited, as when two divers pull at each end of a rope, a buddy-line, when diving in murky waters in order to stay in contact. The tactile contact is an important aspect of social interaction but even more so might the tactile contact be that people avoid, like hitting the other on the nose. These aspects are most probably important when for example building trust between people.

5.3 Importance of haptic feedback in user interfaces

A number of studies have shown that adding haptic force feedback improves single users' performance when manipulating virtual objects. The added value of haptic force feedback lies in peoples' ability to feel the object they manipulate, which makes interaction faster and more precise. It is also found that haptic force feedback increases people's feelings of realism both in the sense that they themselves are actually present in a virtual environment and that objects are actually there to. In some cases it is found that people feel more socially present with other people in shared environments when haptic feedback is provided. Finally people usually report that they feel more confident in and can perform tasks better in haptic virtual environments compared to when only visual feedback is provided.

In one study the task was to put a peg in a hole simulating assembly work (Gupta et al., 1997). It was found that task performance was generally better in the real world than in all the other conditions. As Fitt´s law (1954) predicts, task completion time increased with task index of difficulty, and this was the case in both the real world condition and in the virtual condition with environment sounds, visuals and haptic force feedback. In another set of trials many combinations of modalities were compared such as a real world

condition, 3D visuals/force feedback/sound, 3D visuals/force feedback/no sound, 3D visuals/no force feedback/sound and 2D visuals/force feedback/sound. Haptic feedback was provided by two Phantom force feedback devices. One Phantom timble was positioned on the index finger and thumb respectively on the dominant hand so that the subject could grip an object with two fingers. The results showed that haptic force feedback shortened task completion times significantly compared to the other modes in the virtual environment condition. Gupta et al. (1997) conclude that force feedback provides a superior human-machine interface that fully exploits the hand-eye coordination. There was some evidence that 3D visuals improve the time marginally to move the peg compared to 2D visuals but no significant improvements were found for adding environmental sound such as collision between objects. Another study in which Fitts' law was applied showed that this law did not hold when haptic feedback was not provided in a cube reaching task, but Fitts' law did hold when haptics was added (Mason et al., 2001). In this study however, the haptic condition included real physical objects that were augmented by superimposed graphic visuals of cubes, which means that subjects saw graphical objects but felt real objects in the haptic condition whereas they only saw a graphical object in the non-haptic condition. Subjects took significantly longer time to reach for a graphic only cube than an augmented haptic cube.

Another study (Hasser et al., 1998) showed that the addition of force feedback to a computer mouse improved targeting performance and decreased targeting errors. It has also been shown that if people get haptic force feedback from the context, their performance is improved (Wang and MacKenzie, 2000). Results from that study showed that if a person gets haptic feedback from a table when sliding an object between two targets, performance is better than if moving the object through space without haptic feedback from the context.

The result that haptic feedback improves performance is not without exceptions as some results indicate (Hurmuzlu et al. 1998; Oakley et al., 2000). In one study a non-haptic condition was compared with a condition were haptic feedback was provided with an exoskeletal device that covered the user's arm and tracked the motion of the shoulder and the elbow. One task was to trace a curve on a virtual drawing tablet and several performance measures were obtained such as time needed to complete the tracing, the accuracy of curve tracing and the stability in thickness of the trace. The results showed that the time needed did not differ if a curve was traced in a non-haptic environment compared to if haptic feedback was added. Neither did the accuracy of curve tracing differ between the haptic and the non-haptic environment but the stability in thickness did differ significantly in favour of the haptic condition. Subjects also reported that they felt that it was easier to perform the drawing task with haptic feedback than without. The researchers could not explain why certain task characteristics were harder to perform without haptic feedback

than others but argued that other researchers also have found similar results. Hannaford et al., (1991) stated that the effect of various sensory feedback modes on performance is generally task and system dependent.

In two studies the effect of providing force feedback to people that played melodies with computer-based instruments was examined (O'Modhrain, 2001). It was found in the first study, that haptic feedback only marginally improved performance. In the second study, it was found that haptic feedback in fact to a certain extent confused and misled experienced string players, and that transfer of skills from the real task therefore was not promoted when using a computer-based instrument.

5.4 Haptic interpersonal communication interfaces

A number of projects have explored haptic interpersonal communication interfaces or devices and several interesting aspects of haptic communication have been addressed. Three new ways of communication were explored in the project Feather, Scent and Shaker (Strong, and Gaver, 1996). One of the new ways of communication was a set of shakers. Two people had one shaker each and shaking one device caused the other to vibrate and vice-versa. The persons that communicated in his way did not have any graphical representation of the other person available to them.

In the project inTouch (Brave et al., 1998) the haptic system consisted of two devices that had three cylindrical rollers each. These rollers were linked to each other so that direct continuous physical force on one of the rollers moved one of the rollers on the second device. Two people separated by distance could, passively feel the other person's manipulation of the rollers, cooperatively move the "shared" rollers or fight over the state of the rollers. In another project, PsyBench (Brave et al., 1998), a shared physical workspace allowed distributed users to cooperate by manipulating the same objects physically. However, the users did not share the same graphical interface and did not see a representation of their respective proxy. That meant that it was not possible to see if one person moved his hand towards an object with the intention of picking it up. A representation of a hand is important in order to give another person a cue not to pick up that same object when two persons collaborate in the same interface.

An interesting study was performed in the project HandJive (Fogg et al., 1998), investigating people's need for lightweight and unobtrusive communication and gaming in situations where people have to be quiet like in classrooms. The haptic communication device used in this study was mobile and small enough to fit in the palm of the user's hand. Lessons learnt about haptic

communication was that people tended to compete if the haptic feedback was direct and continuous, implying that indirect haptic feedback and preferably discrete positions induced more cooperative behaviour.

One step further was taken in a project called ComTouch where Chang et al. (2002) investigated how vibrotactile feedback could be used to complement voice communication. A vibrotactile device was developed that could be attached on the back side of a mobile phone. The users pressed the device with one finger in order to send a vibration to the other person while communicating with voice at the same time. It was found that people developed a vibrotactile encoding system similar to Morse code. People also used the vibrotactile feedback in order to emphasize what they were saying by synchronizing their tactile pattern with their speech. Furthermore, it was found that people appeared to use vibrotactile signals to indicate that they intended to speak, as a turn-taking marker. Finally, people mimicked each other's signals and this was argued to function as a way of getting the other person's attention or more symbolically, as patting the other person on the arm.

5.5 Haptic feedback in shared virtual environments

In the real world, haptics is frequently involved in human-human interaction, like hand shaking or tapping someone on the shoulder. Handing over objects is for example a common event in face-to-face interaction. A frequent and watchful example of this occurs when being given a cup of coffee in an airplane – both the flight attendant and the customer have to pay attention to subtle haptic signals to ensure that the hand off is securely accomplished. In relay racing, it is essential that athletes pass a stick from one to another securely when both are in motion in order for them to be successful. The question is how such an event can be supported when the interaction takes place in a shared virtual environment.

According to Gupta et al. (1997) multimodal virtual environments will in the future allow integration of design evaluation techniques with CAD systems in the early design stages and potentially eliminate the need for physical prototypes. Imagine for example a globally distributed group of designers from different disciplines that discuss the integration of each designer's contribution to the construction of a part of a new airplane that is being built. Apart from the fact that they need good quality audio and maybe even video in order to be able to negotiate decisions about the design, they would also benefit from sharing the same haptic object-space. Designers and artists collaborating over distance, may then benefit from being able to jointly manipulate work models, feel the form, weight, surface friction, texture and softness or hardness of objects remotely, or hand off objects to each other in virtual space. However as

yet, no single touch display can provide feedback that is perceived by the user as totally real.

Although not as well studied as single user interface interaction, a few authors have investigated issues regarding joint manipulation of virtual objects in a haptic collaborative virtual environment (Ishii et al., 1994; Basdogan et al., 2000; Sallnäs et al., 2000; Oakley et al. 2001; Sallnäs 2001; Hubbold, 2002; Jordan et al., 2002; Sallnäs and Zhai, 2003; Sallnäs, submitted).

In one study by Basdogan et al. (2000), subjects were asked to play a collaborative game in virtual environments with one of the experimenters who was an "expert" player. The players could feel the objects in the common environment. They were asked to move a ring on a wire in collaboration with each other such that contact between the wire and the ring was minimized or avoided. The two persons participating in the experiment could not communicate with each other verbally.

Results from this study showed that haptic feedback improved task performance when pairs of people worked together. A significant interaction effect was also found, showing that the order of the conditions mattered in the sense that using the visual only system first followed by the visual/haptic system resulted in higher performance, than if the visual/haptic system was used first followed by the visual only system. The authors argue that the initial visual only experience provided training which was significantly enhanced by the addition of haptic feedback, but not vice versa. Another explanation is conceivable, that when subjects had worked with the system with haptics, that generally produced superior performance, they found it harder to only use visual feedback afterwards. On the contrary, if users worked with the visual only system first, which is what people generally know from other computer interfaces, only a positive effect by the haptic feedback was obtained. In this study it was also found that haptic feedback enhanced perceived togetherness significantly. The sense of togetherness was measured by a questionnaire that focused on if the user perceived that they worked with a real human or a computer and in the latter case how realistic the interaction with the other was felt to be. Interestingly females reported a higher level of togetherness than males. The users reported that they thought that they collaborated with a man to a significantly higher extent when they used the visual/haptic system. Togetherness significantly decreased with age and increased with computer use. One has to bear in mind however, that only ten subjects in total participated in the experiment in a within group design.

A study performed in this thesis showed that subjects not only performed tasks significantly faster but also more precisely when manipulating objects together in a haptic compared to a nonhaptic collaborative virtual environment (Sallnäs

et al., 2000; Sallnäs 2001). In this experiment one task required that subjects lifted cubes by pushing from each side of the object in order to build two piles from eight cubes while another task was to build one cube out of the same eight cubes. Two other tasks required that subjects placed cubes in formations on the floor and in the last task subjects navigated, close together, around a formation. It was found that people took significantly longer time to perform the five tasks in the visual only condition without haptic feedback than in the condition with haptic feedback (Sallnäs et al., 2000). It was also found that subjects made significantly more mistakes in performing the two tasks that required that subjects lifted cubes in the visual only than in the visual/haptic condition (Sallnäs 2001).

Three questionnaires were also administered after the experimental sessions that measured social presence, virtual presence and perceived performance. The social presence questionnaire was equivalent to the one used by Short et al. (1976), the presence questionnaire was a slightly modified version of Witmer and Singer's questionnaire (1998) and the performance questionnaire was constructed specifically for the experiment. The social presence questionnaire in contrast to the togetherness questionnaire by Basdogan et al. (2000) focused more on how the social interaction was perceived regarding how personal, social and warm etc. it was and not so much about if the person perceived the co-worker to be real or not. When administering the social presence questionnaire it is taken for granted that persons that participated in the experiment believe that the co-worker is a real person. A large difference compared to the experiment by Basdogan et al. (2000) is that in the experiment presented here (Sallnäs et al., 2000) the participants could talk to each other on a telephone channel.

The results showed that when haptic force feedback was provided subjects' perceived virtual presence was significantly improved but not their social presence (Sallnäs et al., 2000). An explanation for this could be that haptic feedback improved the feeling of realism and control and interactivity, very much in accordance with Katz (1989) who argued that the touch is the primary sense for proof of realness, but that the audio contact over the phone channel is more important for social presence than haptic feedback is. It has been repeatedly shown that providing audio communication is the most important for increasing social behaviour but that for example adding video does not make as conclusive improvements. It was also shown that people reported that they performed the tasks and the collaboration significantly better when getting haptic feedback from the objects, the context and the other person's movements.

Jordan et al. (2002) found that haptic feedback increases co-presence in a study were pairs of people lifted cubes collaboratively. It should be noted however,

that co-presence is in this study measured by the same items that were used in order to measure togetherness in an earlier study (Basdogan et al., 2000). In the study by Jordan et al. the task was for two people to lift a cube together by pushing from each side of the cube and keep the cube off the "ground" for as long as possible (max 2 minutes) once it was lifted. The application was run on a rather fast network between London and Boston. The subjects do not seem to have been able to communicate verbally during the experiment.

In another study the use of haptic cursor communication mechanisms was investigated in a two-dimensional multiuser interface (Oakley et al., 2001). The interface was a synchronous groupware toolkit where users could create and edit text items, rectangular groups, oval groups and links between groups. All items could be freely moved, edited and resized. Telepointers were provided so that all users saw the other as well as their own telepointer. Users worked simultaneously in a work surface that was larger than the computer screen and they could choose to work in a space not visible for the others unless they scrolled towards that part of the work surface. In the haptic condition users could feel resistance when they pushed the cursor into the wall of the window and could in that way scroll in that direction. The users' cursors could be felt and pushed, one user could force another user to go the path that the first user goes, a "locate" tool allowed a person to be guided to another user and finally a "grab" tool allowed one user to force another user to move to her/him. Observations of how people used the haptic features during the performance of a task showed that some pairs did not use the haptic features whereas others used them a lot. When the features were used people usually warned the other verbally about it first before doing anything. The "locate" tool was used more than the "grab" tool, probably because it is more polite to go to someone than to grab someone. Because users in the haptic condition could locate one another they used more of the available space for producing the diagram that was their task. It was also easier for the users to discuss the result when they could locate the other while talking about different parts of the diagram. Results showed that haptic feedback significantly increased presence (ITC Presence Questionnaire), and that the usability was significantly improved in the haptic condition measured by a usability questionnaire. The ITC Presence questionnaire measures the level of spatial presence and engagement and finally naturalness that is how realistic the interaction is perceived to be compared to interaction in the real world. Ratings on items measuring workload showed that the haptic condition was perceived to be significantly more physically demanding. These results show once more that haptic feedback improves perceived presence and perceived usability in a shared virtual environment.

In one project an application with haptic feedback was developed for testing how a stretcher can be carried through a detailed virtual model of a chemical processing plant (Hubbold, 2002). The virtual model had obstacles in the

form of pipework, vessels, pumps, valves other machinery and handrails. The task was for the two carriers to manoeuvre the stretcher past the obstacles to get from the location of a hypothetical accident to an evacuation point, such as a helicopter pad. Two different haptic feedback systems were used. One was an Argonne Arm with six degrees of freedom that was eight-foot long, that had two full scale surrogate handles and that could give output up to 8 lbs force. The second haptic feedback system was a standard desktop Phantom with small surrogate handles attached. The haptic feedback features allowed each user to feel pushing and pulling forces transmitted through the stretcher, collisions between the stretcher and walls or other obstacles and a feeling of sliding along a wall. No empirical tests have been performed yet of this application but observations indicate that it is possible to predict how a stretcher could be carried in the real environment based on the training in the virtual model (Hubbold, 2002). It was found to be extremely difficult to manoeuvre the stretcher without haptic feedback, because the carriers tended to overcompensate when only visual feedback was provided. Haptic feedback was found to be very effective because very subtle haptic cues could be felt when the other person manoeuvered the stretcher and these cues seemed to be used for managing the task. The large difference of scale between the Phantom arm and the Argonne Arm made the interaction unequal in that an aggressive movement by the Phantom user could very easily throw the Argonne Arm user off balance, which also happened during testing.

Another project also investigated interaction in a heterogeneous collaborative virtual environment (McLaughlin et al., 2003) in which one person got haptic feedback from a Phantom whereas the co-worker got tactile feedback on each finger of the hand from a CyberGrasp. The task was for the subject using a Phantom to communicate information through tapping "Morse code" at the fingers of the subject using a CyberGrasp. What makes this project interesting is that the architecture developed supports multiple users, with heterogeneous haptic devices, over a non-dedicated channel which adapts to network delays. However, no empirical comparisons were made between the haptic and a nonhaptic condition or between the haptic and a real world condition.

Intuitively haptics may play a critical role when people pass objects between each other. The giver has to sense that the recipient has firmly grasped the object before releasing it. The recipient has to feel that the giver is releasing it before taking it towards oneself. It is difficult to imagine that such a task could be accomplished without haptic feedback. A study was performed in this thesis in order to investigate how haptic force feedback affects people's performance when handing over objects in a collaborative virtual environment (Sallnäs and Zhai, 2003; Sallnäs, submitted). In an experiment, subjects passed a series of cubic objects to each other and tapped them at target areas. Their performance with and without haptic force feedback was evaluated. The subjects could not

communicate verbally with each other during this experiment. Furthermore, the study was placed in the framework of Fitts' law and it was hypothesized that object hand off constitutes a collaboratively performed Fitts' law task, with target distance to target size ratio as a fundamental performance determinant. Our results showed that task completion time indeed linearly increased with Fitts' index of difficulty, both with and without force feedback. It was a little surprising that the time required for handing over objects did not differ significantly between the haptic and nonhaptic condition even though there was a large difference in favour of the haptic condition. However, the error rate was significantly lower with haptic feedback than without.

Furthermore it was found that people perceived virtual presence and social presence were significantly higher when they collaborated in the visual/haptic virtual condition than in the visual only condition (Sallnäs, submitted). It was also found that haptic feedback significantly increased perceived performance when people performed a Fitts' law tapping task collaboratively. The result regarding social presence in the previous study (Sallnäs, submitted) and the results obtained by Basdogan et al. (2000) and by Jordan et al. (2002) that got similar results can be compared to the results by Sallnäs et al. (2000) where no difference was found between a haptic and a nonhaptic condition. These results analysed together suggest that when a telephone channel is provided the effect of haptic feedback on perceived social presence is overshadowed.

6 Summaries

6.1 Article A

Sallnäs, E-L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction*, *7*(4), 461-476.

6.1.1 Aims and background

Haptic feedback is a natural ingredient in face-to-face interaction between people, and it also serves important functions for communication. Examples are handshakes and a pat on the back that are powerful communicative and symbolic events that convey information about relations, status and emotional states. It is also intuitive for people to combine gestures, deictic references and joint manipulation in collaborative environments. An experimental study was performed in order to test the hypotheses that a shared virtual environment supporting the touch modality will increase social presence, perceived virtual presence, improve task performance and increase perceived task performance.

6.1.2 Method

A between-group design was used and the independent variable in this experiment was the interface conditions with two treatments: a haptic and a nonhaptic virtual environment. Twenty-eight subjects participated in the experiment. Of these subjects, 14 were men and 14 were women. The subjects,

in different locations, performed five collaborative tasks in both conditions. The haptic devices used in the tests were two 1.0 Phantoms (Figure 6). In the first condition that included haptic force feedback, the subjects obtained haptic force feedback from the dynamic objects, the static walls and the other person in the shared virtual environment. The subjects could simultaneously manipulate the dynamic objects that were modelled to simulate real cubes with form, mass, damping and surface friction. The subjects could also hold on to each other by pushing a small button on the haptic device. In the second condition the subjects had no haptic force feedback and could not hold on to each other. The haptic device then functioned as a 3D-mouse. Voice communication in both conditions was provided through a telephone connection using headsets.

Figure 6. Two persons are doing the tasks together in a shared virtual environment. Users see the same view on their screens and simple avatars represents them in the form of a green or a blue sphere. They hold the Phantom pen with their dominant hand, the person on the left is right-handed whereas the other person is left-handed.

Task performance was measured by the total time it took the pairs of subjects to perform the five tasks. The data was obtained through analysis of video recordings of the experimental sessions. The subjects' perceived task performance, perceived virtual presence and perceived social presence were measured by questionnaires.

6.1.3 Findings and conclusions

The results showed that haptic force feedback significantly improved task performance, which means that the tasks were completed in less time in the haptic force feedback condition. Subjects used an average of 24 minutes to perform five tasks in the haptic force feedback condition as against 35 minutes in the condition with no haptic force feedback. The questionnaire that measured

perceived performance showed that the subjects in the haptic force feedback condition perceived themselves to be performing the tasks significantly better. The analysis of data from the virtual presence questionnaire showed that haptic force feedback added significantly to people's perceived virtual presence compared to the condition where haptic feedback was not provided. However, the analysis of people's ratings on items measuring social presence showed that the conditions did not differ significantly. This means that people did not feel that they were more socially present in the haptic environment.

6.2 Article B

Sallnäs, E-L. (2001). Improved precision in mediated collaborative manipulation of objects by haptic force feedback. In G. Goos, J. Hartmanis and J. van Leeuwen (Series Eds.) and S. Brewster and R. Murray-Smith (Vol. Eds.), *Lecture Notes in Computer Science: Vol. 2058. Haptic Human-Computer Interaction* (pp. 69-75). Heidelberg, Germany: Springer.

6.2.1 Aims and background

The aim of this study was to analyse data collected in the experiment presented in article A in order to investigate if haptic force feedback improved the precision with which subjects could coordinate their manipulation of objects in the shared virtual environment. The frequency of failure to lift cubes together was analysed as a measure of precision. The results from the analysis in this paper show in more detail how performance is affected by haptic force feedback for joint manipulation of virtual objects.

6.2.2 Method

Data was analysed for the twenty-eight subjects that participated in the experiment presented in article A. The method, the design of the study, and the task has been presented previously in this thesis in the summary of article A. As mentioned the collaborative desktop virtual interface was the independent variable in the experiment and there were two conditions, one haptic interface and one nonhaptic interface. The haptic environment consisted of a room with constraining walls, ceiling and floor and it contained eight dynamic cubes that initially were placed on the floor. In the haptic environment subjects could feel the shape, weight, friction and softness of the objects and the other persons representation in form of a sphere as well as the forces applied by the other person. In the version without haptic force feedback the Pantom functioned solely as a 3D mouse, as the user could feel neither the cubes, nor the walls, nor the other user in the environment. The subjects could lift the cubes in two

different ways. Either the users collaborated in lifting the cubes by pressing into the cube from opposite sides and lifting upwards simultaneously, or a single user lifted a cube by pressing it against the wall and pushing it upwards. For the analysis in this paper, data on the frequency of failures to lift the cubes collaboratively were collected for two of the five tasks that the subjects performed. These were task A and task C which both required subjects to lift cubes in order to complete the task. Task A consisted of lifting eight cubes together in order to build one cube and task C consisted of lifting eight cubes together in order to build two piles. The video recordings were analysed in order to collect the frequency of failures to lift the cubes collaboratively as a measure of precision in task performance. The operational definition of failure to lift a cube were, that two subjects positioned their representations on each side of a cube and tried to lift it, but failed to lift or transport the cube in order to proceed one step in performing the task.

6.2.3 Findings and conclusions

Frequencies of failures to lift cubes together were analysed with ANOVA (analysis of variance). Results showed that there was a significant difference between conditions regarding subjects´ ability to lift cubes both in task A and in task C. The results in the experiment presented in article A suggested that it took significantly longer time to perform tasks in the condition without haptic force feedback. In the earlier analysis subjects also judged their performance as significantly better in the haptic environment. The analysis that is presented in this paper show that it is significantly more difficult to coordinate actions with the aim of lifting objects in a three-dimensional desktop virtual environment without haptic force feedback. These results suggest that a major part of the difference regarding time to perform tasks can be explained by the fact that subjects' precision when lifting cubes without haptic force feedback is significantly lower. It should be noted that even in the haptic condition manipulation of virtual cubes was not effortless and subjects did fail a number of times even with haptic force feedback. But subjects performed the actions that they had planned more consistently and they did not shift strategy in the collaborative task as often because of failure to lift a cube.

6.3 Article C

Sallnäs, E-L. (conditionally accepted). Effects of communication mode on social presence, presence and performance in collaborative virtual environments. *Journal of Presence: Teleoperators and Virtual Environments*.

6.3.1 Aims and background

In most virtual worlds the means by which people communicate is text-chat. An experimental study was conducted in order to investigate to what extent collaboration in a collaborative virtual environment (CVE) is affected by an audio communication channel or video connection in comparison to text-chat. The aim of this experiment was to investigate to what extent the medium affects the notion of social presence, presence and performance in a CVE. Another aim was to study to what extent the medium affects the collaborators' communication behaviour regarding the time to complete the task in a CVE, frequency of words spoken and finally frequency of words spoken per second. A follow up experiment was conducted with the aim of comparing the results in the first experiment with results from collaboration in a web-environment. Subjects communicated either by telephone or a video conference system in the follow up experiment. Subjects performed the same task in the web environment and the information content was also the same as in the CVE. In the analysis, the results from the interaction in the CVE audio and CVE video conditions in the first experiment were compared to results from the interaction in the web audio and web video conditions in the second experiment.

6.3.2 Method

A between-group design was used and in total eighty subjects participated in the two experiments. They were assigned to pairs, with one woman and one man in each pair, and they performed a decision-making task together. In the first experiment a CVE with three conditions, video conference, voice and text-chat was studied. The data was analysed using one-way ANOVA. In the follow-up experiment a two-way ANOVA was performed on the data from the CVE audio and CVE video condition from the first experiment and data from the web audio and web video condition from the second experiment. Dependent subjective variables in the experiments were perceived social presence, perceived presence and perceived task performance. These were measured by questionnaires. Furthermore, the dependent objective variables pertaining to the interaction were time to finish the task, frequency of words used in communication, and words spoken per second. The dialogues were transcribed from the video recordings. Two PowerBook PCs networked via Ethernet were used in both experiments. In the voice condition, two telephones

with headsets were used. In the condition with a video connection, two 21-inch television monitors were used plus two telephones with headsets providing the voice communication. In the text condition the subjects communicated via the text-chat that is a feature in ActiveWorlds. The CVE was constructed in Active Worlds and the environment consisted of an exhibition with information points that included posters and QuickTime movie clips with audio (see Figure 7). Human-like avatars represented the subjects.

Figure 7. The exhibition in the ActiveWorlds environment with posters and QuickTime movie clips with audio.

Figure 8. The web environment with posters and QuickTime movie clips with audio.

The web-environment (see Figure 8) was designed as a web-site with the same information points (posters and QuickTime movie clips with audio) as in the CVE. In the web-conditions, the information points were placed beside each

other in one web page in a sequential fashion. The subjects had no avatars in the web-conditions and they could not see the other subject's cursor.

6.3.3 Findings and conclusions

The results from the one way ANOVA showed that text-chat is very different compared to the audio and video communication media. When communicating in a CVE with text-chat, people perceived their presence and social presence to be significantly lower compared to when they interacted in a CVE with audio or in a CVE with video conference. Making joint decisions in the CVE text-chat condition took significantly longer time, about 29 minutes, and dialogues were significantly scarcer in text-chat than in both CVE audio and CVE video. The result that significantly fewer words per second were used in text-chat than in both CVE audio and CVE video suggest that the tempo is very slow and indicates indirectly that there are more pauses in the dialogue.

The results regarding audio and video conditions are somewhat more complicated to interpret. There are no differences regarding perceived presence or social presence between CVE video or CVE audio conditions when analysed in a one-way ANOVA. However, when the CVE and web conditions and communication media are compared in a two-way ANOVA a significant difference is found regarding perceived presence depending on if audio or video communication is used. People generally perceive that they are more virtually present in video conditions than in audio conditions. Interestingly, a significant interaction effect shows that this difference is very strong in web environments but not at all as important in CVE environments.

Even though differences are not significant, mean values show that collaborating in CVE audio was quite fast, about 10 minutes, whereas people spent more time in CVE video, about 16 minutes. Furthermore, people had slightly less extensive dialogues in the CVE audio (mean =1084 words) condition than in the CVE video (mean =1243 words) condition. The difference between CVE audio and CVE video becomes clearer with the finding that significantly fewer words per second are used in CVE video (mean = 1.3) than in CVE audio condition (mean = 1.9). This finding shows that the tempo is slower in the CVE video condition and indirectly shows that there are more pauses. This suggests that the video condition is in fact different from the audio condition.

Results from the two-way ANOVA in the follow-up study showed that it took significantly less time to complete the task in the two audio than in the two video conditions but no significant difference was found between the CVE and web conditions. However, there was a significant difference in the number of words used per second between the CVE conditions and the web conditions. If mean values are compared a high amount of words per second are used in

the web audio (mean = 2.2) and web video condition (mean = 2.0) and a little less are used in the CVE audio condition (mean = 1.9), whereas much fewer words per second are used in the CVE video condition (mean = 1.3). This means that the tempo in the dialogue is slower in the CVE video condition, that people spend a long time in it and that people have the most extensive dialogue there, compared to the other conditions. A factor analysis showed that the questionnaire items measuring presence, social presence and perceived performance in this study did measure three different phenomena even though some of the presence items overlapped with the perceived performance items.

6.4 Article D

Sallnäs, E-L., & Zhai, S. (2003). Collaboration meets Fitts' law: Passing virtual objects with and without haptic force feedback. In M. Rauterberg, M. Menozzi & J. Wesson (Eds.), *Proceedings of INTERACT´03* (pp. 97-104). Amsterdam: IOS Press.

6.4.1 Aims and background

In this study the focus is on object hand off as a paradigm of evaluation, because it is a type of joint haptic event between two people that requires coordinated action to accomplish. For example, when being given a cup of coffee in an airplane – both the flight attendant and the customer have to pay attention to subtle haptic signals to ensure that the hand off is securely accomplished. It was hypothesized that object hand off constituted a collaboratively performed Fitts' law task, with target distance to target size ratio as a fundamental performance determinant. Fitts' law has traditionally been used as a model for performance of individual tasks like target pointing. According to Fitts' law, for the same distance, the greater the object is, the faster one can point at it. When handing over an object, such a relation may also hold since the larger the object is, the more relative tolerance it allows in the hand off process in terms of where one could hold the object. Given this background, the aim was to study two issues in a collaboratively performed hand off task: Fitts' law applicability as a task performance model and the role of haptic force feedback.

6.4.2 Method

A within group design was used in this experiment and twenty-two subjects participated in it. Each subject was seated in front of a haptic display system in separate rooms. The haptic and the nonhaptic virtual environment were implemented using Reachin Technologies AB's API on a Windows 2000 PC. The haptic display systems used in this project consisted of two displays from

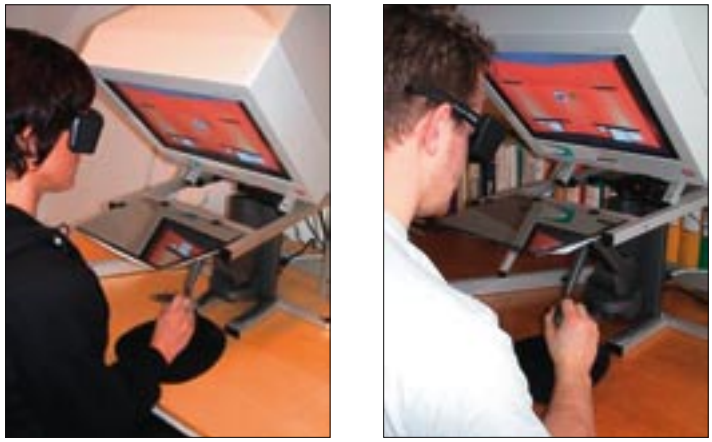

Figure 9. Two persons collaborating in the virtual environment using one ReachIn Display system each. They hold a desktop Phantom pen each in order to grasp an object and pass it between each other.

Reachin Technologies AB with two Desktop Phantom force feedback devices from SensAble Technologies, Inc. (Figure 9).

The subjects were not able to communicate verbally during the experiment neither by voice nor by text. The three-dimensional haptic collaborative interface was designed as a room with two larger shelves, on top of which six cubes were placed, three on each side (Figure 10). The room also contained two smaller shelves that served as target areas, underneath the two larger shelves. Two cursors, coloured green and blue, corresponded to the tip positions of the two Phantom probes. To begin the task, subjects were instructed to alternately grasp a cube, lift it and hand it to the other subject who tapped the second target shelf with the cube. Subsequently the second subject returns the cube to the first subject who then proceeds to tap the first shelf and so on. Task difficulty was manipulated by changing cube sizes in randomised order. The distance (D) between the target shelves and the size of the target shelves were fixed.

In both the haptic and the nonhaptic environment it was possible to grasp a cube by touching the cube with a cursor and then pressing the button on the haptic device. The haptic user interface was developed so that all surfaces in the environment were touchable and thus provided haptic force feedback. When two people both held an object (both buttons pressed down), they could feel each other's actions such as pushing or pulling, which facilitates haptic information exchange between the two persons during the hand off operation. In the condition without haptic force feedback, the Phantom functioned solely as a 3D mouse without force feedback.

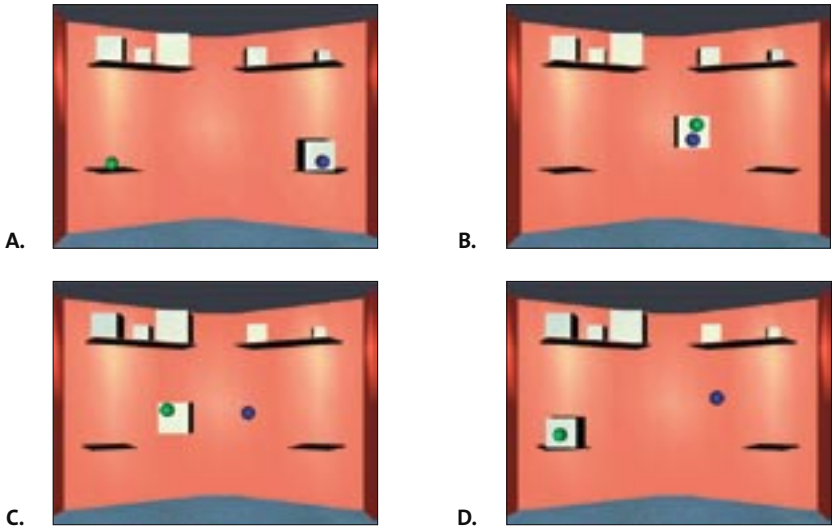

Figure 10. A sequence showing two subjects (green and blue sphere) performing a hand off in the collaborative virtual environment: **a** one person grasps a cube **b** that person passes the cube to the second person **c** the second person moves the cube towards the target area **d** and taps the cube at the target shelf.

6.4.3 Findings and conclusions

The results clearly demonstrated that the time to accomplish a hand off task depended on Fitts' index of difficulty. As the size of the object that was handed over decreased, the time to complete a hand off trial successfully increased logarithmically. The correlation between the trial completion time, which combines multiple stages and both individuals' actions, and Fitts´ law ID was higher than expected. In the haptic condition the correlation metric r^2 was 0.98, and in the nonhaptic condition r^2 was 0.93. Results showed that the average trial completion time to perform the hand off task successfully, did not differ significantly between the haptic and nonhaptic conditions. Results showed however, that subjects dropped significantly more cubes in the nonhaptic than in the haptic condition.

In conclusion, this study proved that Fitts' law could be used as a performance model for a new class of complex tasks, collaboratively performed by two individuals, both in the haptic and non-haptic condition. It also showed that haptic feedback could help participants' to coordinate their actions and reduce failures significantly in the case of handing over objects in a shared virtual environment.

6.5 Article E

Sallnäs, E-L. (submitted). Passing virtual objects collaboratively with and without haptic feedback: effects on social presence, virtual presence and perceived performance.

6.5.1 Aims and background

This article builds on the experiment presented in article D where Fitts' law was applied to a collaboratively performed hand off of a graphical object. In this article however, the focus is on subject's ratings in order to investigate the effects of haptic feedback on subjects' perceived social presence, perceived virtual presence and perceived performance when passing an object to each other in a shared virtual environment. In article A an experiment was presented where it was found that haptic feedback did not increase subjects' perceived social presence significantly, but indeed increased their perceived virtual presence and perceived performance significantly. In the experiment presented in this article (E) subjects did not have audio communication that might have overshadowed the effects of haptic feedback on perceived social presence in the earlier study (article A) and therefore results from the two studies were compared.

6.5.2 Method

Ratings from eighteen of the subjects from the experiment presented in article D were used for the analysis in this study. Subjects collaborated in nine pairs, each consisting of one woman and one man except one pair with two men. The independent variable was the shared virtual environment with two conditions, a haptic condition and a nonhaptic condition. The dependent variables were the subjects' ratings on three questionnaires measuring social presence, virtual presence and perceived performance. Perceived task performance was measured by a questionnaire with 14 items, using bipolar Likert-type seven-point scales. The questionnaire focused on the users' evaluation of their own task performance when using the system, how well they understood the system and to what degree they felt that they learned how to use the system and also their skill level in using specific features in the system. The social presence questionnaire was inspired by the Social presence theory (Short et al., 1976). The definition of social presence in this experimental study was "feeling that one is present with another person in a shared environment". A bipolar seven point Likert-type scale was used. The questionnaire consisted of 34 questions. The first eight questions measure how personal versus impersonal, warm versus cold, sensitive versus insensitive, sociable versus unsociable etc. subjects perceived that the social interaction was. The other social presence items measure

to what extent persons feel that they are socially present, that they understand the other person's intentions, emotions and reactions and to what extent they feel that their reactions are perceived by the other person. In this experimental study presence defined as, "feeling as if being in a mediated environment" was referred to as virtual presence. Virtual presence was measured using Witmer and Singer's (1998) Presence questionnaire (PQ) with Likert-type seven-point scales. The questionnaire consisted of 30 questions. Aspects measured are the feeling of being able to control events, whether the virtual environment seems responsive to actions and how involved the subject became in the experience. Furthermore, questions about how natural the interaction felt is included, how consistent the virtual environment was with reality, and how natural the movement control was. Finally, aspects measured are whether the controls or the display interfere or distract from task performance, and the extent to which they felt able to concentrate on the task.

6.5.3 Findings and conclusions

Results from this study showed that haptic feedback increased subject's perceived virtual presence significantly. This shows that haptic feedback makes the impression that one is actually there in a mediated environment much stronger than if one does not have haptic feedback. There was also a significant difference in how well subjects perceived that they performed in the haptic virtual environment compared to a nonhaptic environment. Subjects felt more secure and able to pass cubes to each other in the haptic environment. In this study results also showed that haptic feedback increased perceived social presence which means that people to a larger extent perceived that they were present with another person in a mediated environment. Following from the topics in the social presence questionnaire people perceived that the interaction was more personal, social and that they understood and could express feelings and intentions to a larger extent and that they perceived the other person as more real and socially present. It was found that perceived performance correlated significantly with virtual presence but not with social presence even though it was close in the haptic condition and that social presence correlated significantly with virtual presence. This result indicates that virtual presence varied with the notion that people had about how they performed the task and even that virtual presence might predict performance to a certain extent. Social presence varied with virtual presence but did not vary as strongly with perceived performance which might suggest that even a situation in which pairs feel that they perform poorly, they can in fact still feel a high sense of social presence which is in fact not so hard to imagine. Poor performance might lead to frustration, which might affect the emotional state between the collaborators, that could lead to an increased notion of social presence. When comparing results from the study presented in article A, and the study

presented here, the analysis of the eight questions that were the same in both studies showed that when voice communication was provided, haptic feedback did not improve perceived social presence significantly, but that it did when voice communication was not provided. One could argue that this difference might be the effect of the voice communication but it is hard to know how big the effect of the differing tasks and designs are. The results indicate, even though it has to be investigated further, that haptic feedback does increase virtual presence regardless of verbal communication but that social presence depends more on the communication channel used than on haptic feedback.

7 Discussion and conclusions

In this thesis social psychological aspects of collaboration in multimodal virtual environments have been described and evaluated. A basic assumption for the research in this thesis has been that the more modalities used the more rich the interaction is perceived to be and the better people accomplish their goals and their tasks, especially if the task is ambiguous or socially demanding. Most effort in this thesis has been on varying the modalities used and not as much on varying the characteristics of the task performed. A number of measures have been used in order to be able to compare different settings. A social presence questionnaire was developed that was inspired by the social presence theory. A questionnaire measuring presence was used, developed by Witmer and Singer (1998). In this thesis the concept virtual presence was used instead of presence in some studies in order to make the difference clearer in comparison to social presence. A questionnaire measuring perceived performance was developed and used in all experiments. Finally, task performance has been measured in all experiments. The major emphasis is in this thesis put on the comparisons within each experiment, as the possibility to make comparisons across results from different experiments is limited. The main empirical conclusions presented in this thesis are that haptic feedback is advantageous for collaboration in shared virtual environments in different ways and that communication media play an important role for collaboration in shared virtual environments.

7.1 Effects of haptic feedback on joint task performance

The main conclusion that can be drawn from the studies presented in articles B and D in this thesis is that precision in coordinating joint manipulation of virtual objects is significantly improved by haptic force feedback. In article B two people coordinate the forces by which they push from each side of a cube in a shared virtual environment in order to lift it. In the haptic environment the co-workers can feel pushing forces through the object and they feel the texture of the object and can estimate how hard they can push without slipping off the object surface. In article D people coordinate when to grasp and when to release an object alternatively and to tap at a target area without dropping the cube. In the haptic environment the giver senses that the recipient has firmly grasped the object before she releases it, and the recipient feels that the giver releases the object before taking it. Collision forces can also be detected between the virtual object and the shelf or the other person. However, in the nonhaptic environments precision meant using vision only, in order to align a graphical object and the graphical representation of a person like for example when choosing a function via an icon in a three dimensional graphical user interface. It could be argued that tasks in these experiments are tailored to be performed better in a haptic virtual environment, but most tasks involving joint manipulation are probably much facilitated by haptic feedback. It is in fact hard to imagine a manipulation task that would not be helped by haptic feedback. We are using graphical interfaces today such as MS Windows to a large extent that are mostly only visual. Even in these interfaces, precision might possibly be improved by haptic feedback although that is an interface that was not designed with haptic feedback in mind. We are just not used to thinking in terms of using touch feedback in computers, neither as single users nor in shared interfaces.

The time spent by collaborators to finish a task has been used as a basic measure of efficiency and it is assumed that there is a relationship between the activities of collaboration and the speed at which a pair performs a task. Results in this thesis show in the case of the study presented in article A that haptic feedback shortens the task completion time significantly, compared to a nonhaptic condition. However, results from the study presented in article D show no significant difference between the haptic and the nonhaptic condition regarding the time spent to complete the task together. This result might have to do with the fact that the experimental task in the study presented in article D provided only short haptic events. In the study presented in article A continuous haptic feedback was essential for improved performance. In that experiment, subjects pushed from both sides of objects in order to move them together, experiencing haptic force feedback during a high proportion of the trial time.

Another reason might be that the nonhaptic condition in the study presented in article D was relaxed from its real world counterpart by the "elastic" property between the object and the cursor. This could have reduced the stringent timing coordination between the two participants and could indeed be taken advantage of in virtual environment design. Even though the result in the study presented in article D did not show a significant difference regarding time it was close (significant at a 10% level) which might mean that more subjects in fact would have shown that haptic feedback shortens completion time also for this kind of task.

In this thesis it was concluded that Fitts' law is applicable as a task performance model for a collaboratively performed hand off task. Fitts' law has traditionally been used as a model for performance of individual tasks such as target pointing. It has played an important role for user interface design and systematic evaluations of different input devices (Card et al., 1978; MacKenzie et al., 1991; MacKenzie, 1992; ISO, 2000; Zhai, 2002). In article D it was hypothesised that although much more complex than simple target tapping, Fitts' law may also hold when passing an object to another person. The larger the object is, the more relative tolerance it allows in the hand off process in terms of the accuracy the giver has to target at the receiver's hand, and the accuracy with which the receiver could "grasp" and hold the object. The results in this thesis clearly demonstrated that the time to accomplish a hand off task depended on Fitts' index of difficulty. As the size of the hand off object decreased, the time to successfully complete a hand off trial increased logarithmically. The correlation between the trial completion time, which combines multiple stages during the task and both individuals' actions, and Fitts´ law ID was higher than expected. In conclusion, it was found that Fitts law's applicability and robustness to this new type of task was very good. This was especially true when haptic force feedback was provided as the correlation was comparable to or better than the goodness of fit in individual target reaching tasks reported in the literature.

7.2 Social presence in haptic collaborative virtual environments

One of the most important questions that have motivated the research in this thesis is how the framework of social presence theory can be applied in evaluating collaborative virtual environments. Earlier research on telecommunication media have suggested that the richer the media used, in terms of the extent of verbal and nonverbal information transmitted, the more social presence is experienced in human-human interaction. Haptic feedback systems now make it possible not only to use vision and hearing but also the

touch modality when interacting in virtual environments. One of the questions that have been investigated in this thesis is whether haptic feedback makes human-human interaction richer in the way social presence theory predicts. Extending this question further, is haptic feedback transmitting cues that are communicative?

The conclusions in this thesis regarding the extent that haptic feedback increases the sense of social presence are mixed. In the study presented in article A people did not report higher levels of social presence after using a collaborative environment with haptic feedback than people that used the nonhaptic environment. In this environment all pairs of people communicated by a telephone channel using headsets, that provide very high quality audio. In the study presented in article E on the other hand, people did report that they experienced a significantly higher level of social presence in the haptic condition than they did in the nonhaptic condition. In the study presented in article E, the collaborating pairs could not communicate verbally. The study in article A was a between subjects design whereas the study in article E was a within subjects design which makes a comparison between the studies difficult. Nevertheless, the conclusion is that haptic feedback sometimes increases the experience of social presence and sometimes does not and that it might be that in the case were it does not, voice communication might overshadow the effects of haptic feedback. It would not be far fetched to suggest, drawing on research that consistently finds that voice makes the biggest difference for social presence, that this same effect is at work in the case of haptic feedback. Haptic feedback is important for perceived social presence but voice communication is probably much more important. However, results from informal interviews after the experiments can provide preliminary indications that haptic feedback can convey communicative cues. Subjects for example commented that:

"You signal (to the other person) that you are at it (the object) when you push from the front (of the cube) because you felt that yourself (that you are on the object) so to say…"

The fact that social presence was perceived to be increased in the haptic environment where verbal communication was not possible, suggests that some kind of information was communicated through the touch modality. Together with the comments afterwards, it seems that intentions could be at least perceived to be communicated, even if it is hard to know if this "signalling" was interpreted in the way intended by the sender or not. On a lower level the intended coordination of action probably is communicated and the forcefulness with which a person handles objects and interacts with the other person could probably communicate emotion. The question if a person can discriminate between whether a forceful push is due to eagerness, playfulness

and liveliness or aggressiveness and hostility, cannot be answered within the scope of this thesis.

7.3 The effect of haptic feedback on perceived virtual presence

It is clear from the two studies presented in article A and E in this thesis, that haptic feedback does increase peoples' sensation of virtual presence significantly. In contrast to the results regarding social presence peoples' perceived virtual presence seems not to be affected by if verbal communication is possible between collaborators or not. This strengthens the idea that these two constructs have indeed different meaning to people and measure two separate phenomena. This was further established by the results from the factor analysis performed in the study presented in article C.

It thus seems as if haptic feedback makes the impression that one is actually there in a mediated environment much stronger than if one does not have haptic feedback. The operationalisation of virtual presence applied to haptic feedback means that people feel that their interaction in the shared environment is more natural and realistic, that they have more control when interacting and that they become more immersed and engaged in the haptic than in the nonhaptic virtual environments. That the touch modality is important for the sense of reality and that people rely very much on haptic feedback for manipulating objects might not be so surprising as already David Katz (1989) pointed out. However, one has to remember that the haptic feedback provided through the system is only from one single point of interaction. This means that even if the haptic interaction has high resolution, it is still far from the touch sensation we get in real life. Therefore it is interesting that this kind of simplified haptic perception makes the experience of realness in a virtual environment so much stronger.

7.4 Perceived task performance with or without haptic feedback

In this thesis the conclusion is, based on the results from the studies presented in article A and E, that people rate their performance to be significantly better when they get haptic feedback from the shared virtual environment. This result is consistent with the results regarding the actual joint task performance that is significantly improved in terms of precision in both studies and in terms of completion time in the study presented in article A. The fact that both perceived task performance and actual task performance are improved is promising

as the fact that the haptic system is new to the people participating in the experiments might make them less confident in using it. Subjects commented, after experimental sessions in this study, on the fact that they felt more secure in handing over cubes in the haptic environment:

"In the environment where you can feel, then you feel what the other person does. If both are at the object, then you can adjust so that both persons help each other to move in one direction. But in the other environment you have no idea what the other person actually does. Then it can happen that you pull in different directions…"

One aim that is often advanced in virtual reality research is to be able to predict task performance through a questionnaire measuring virtual presence. In this thesis an analysis of the extent that actual performance and virtual presence co-vary has not been performed. However, in the case of adding haptic force feedback in the two studies presented in article A, B, C and D both the precision in joint manipulation of objects and perceived performance were significantly increased as well as the experience of virtual presence. Social presence theory to a certain extent also aims at being able to predict performance. In the studies investigating the effects of haptic feedback in this thesis, the results show that such a relation is not as clear. That relationship is probably made more complex by the intervening variable of communication media. This is a question that has to be investigated further in the future.

7.5 Media effects on communication in virtual environments

In this thesis it was found that communication media play an important role when people interact in a shared virtual environment. The most common communication medium used today in virtual environments is text-chat, as in for example ActiveWorlds. In that perspective the finding that communicating with text-chat was found to be significantly different in many respects than voice and video/voice communication is important. It was found that text-chat affected the performance of the joint task in all measured respects compared to the other conditions. Solving the task took significantly longer time, the dialogue was significantly less extensive and the pace of the conversation was significantly slower. These results might not be so surprising, as words had to be typed in the text-chat condition instead of being spoken as in the video/voice and voice conditions. It is striking, however, how large the difference is. Even though text-chat proved to be slow and crude, all pairs managed to complete the task with text-chat with no apparent problem. However there was one instance of flaming were the persons misunderstood each other when communicating

with text and in the end insulted each other. This was handled afterwards in a debriefing where the reason for the misunderstanding, one person taking a long time to answer in the text-chat, made the persons understand why the situation had arisen, and they parted as friends.

A lot of effort has been put into research to investigate the extent that video conference adds benefits to distributed collaboration compared to leaner media. The general agreement is that it is hard to establish that video does make a difference especially regarding task performance. In this thesis in many ways that same conclusion was made, video did for example not make any difference for the amount of words spoken in the experiment presented in article D. But a more complex picture emerged from the analysis of the task performance measures time and words spoken per unit of time. That the voice only and video conference did in fact impact collaborators differently, became clear with the result that significantly more words per time unit were used in the voice compared to the video/voice condition in shared virtual environments. When data from the voice and video/voice virtual reality conditions was compared to data from a voice web and a video/voice web condition, yet another difference in the amount of words used per time unit was found. People used significantly more words per time unit in the web conditions than in the shared virtual reality conditions. If mean values are compared they show that video/voice in virtual reality environments might have unique qualities compared to the other conditions. Most words per second were used in the web audio (mean = 2.2) and web video condition (mean = 2.0) and fewer were used in the CVE audio condition (mean = 1.9), whereas much fewer words per second were used in the CVE video condition (mean = 1.3). The combined results show that the pace of the conversation was significantly higher in the web conditions than in shared virtual environments and was, in the first study, significantly higher in audio than in video. The interpretation suggested is that the amount of words used per time unit indirectly shows the difference in tempo in the dialogue and possibly the amount of pauses in talk during collaboration. It was informally observed that people seemed to tolerate pauses better in the video/voice than in the voice condition when both persons investigated information in the shared virtual environment.

When the task completion times in the voice and video/voice virtual reality conditions were compared to task completion times in the voice web and a video/voice web condition, it was found that it took significantly less time to complete the task in the audio than in the video conditions, but whether a shared virtual reality or a web condition was used did not matter.

Overall, these results show that people, apart from the text-chat condition, used more time in the video condition in total, had a slightly more extensive dialogue, and communicated in quite a slow tempo which might indicate that

the video conference possibly was a more relaxed medium, whereas the audio condition was the most efficient for this task. Finally, it could be argued that the effect of communication media on the tempo of the conversation was most prominent in virtual reality environments.

7.6 Effects of communication media on social presence, presence, and perceived performance

In the next set of conclusions the effects of communication media on peoples' experience of social presence, virtual presence and perceived performance in shared virtual environments are discussed. Again as stated earlier text-chat is today the most common way of communication in shared virtual environments and therefore only a few researchers have investigated the effects of different communication media experimentally. Traditionally social presence has been investigated experimentally but not so much in combination with a shared virtual environment. On the other hand virtual presence, or the more commonly used term presence, has been evaluated mostly in single user virtual environments. However, some researchers in the research area of virtual reality have introduced the constructs of co-presence. Co-presence questionnaires put most emphasis on how real the other person is perceived to be. The social presence perspective that is advanced in this thesis focuses more on how social, personal and emotional the interaction with a remote person is perceived to be.

In this thesis a social presence questionnaire and a questionnaire measuring perceived performance have been developed and evaluated. It was shown in article D that the internal consistency measure of reliability was good for social presence, virtual presence and perceived task performance. A principal component analysis showed that the items in the social presence, virtual presence and perceived task performance questionnaires measured three constructs. The factor analysis on these items showed that the items in the social presence questionnaire measure a different construct than do the virtual presence and the perceived performance items. Furthermore it was found that there was an overlap between the items measuring perceived performance and a small number of virtual presence items, but that the virtual presence items, apart from that, measure a separate construct.

One conclusion in this thesis is that people experience a significant increase in virtual presence when voice or video/ voice communication is added compared to text-chat in a shared virtual environment. This suggests that people feel that their interaction in the shared environment is more natural and realistic, that they have more control when interacting and that the become more immersed

and engaged in the virtual environment when having voice or video/voice communication. When data from the voice and video/voice virtual reality conditions was compared to data from a voice web and a video/voice web condition, a more complex picture emerged. Results showed a significant increase in perceived virtual presence in the two video conditions than in the two audio conditions and a significant interaction effect showed that there was no effect of medium in the virtual reality conditions, but in the case of interaction in a web environment, virtual presence was perceived to be increased by video conference. Why is this the case? No definite answer can be given to this question but a number of reasons can be suggested. It might be that because the shared virtual environment is perceived as a place where one navigates, positions one's avatar in relation to objects and other people it might be that people either are so focused on managing the interaction with the interface that they do not use the video image as much. Another reason might be that the place-like qualities of the virtual reality add many realness cues and that these even out the effect of the video channel. The complexity also lies in the fact that only collaborative environments have been compared and that the virtual presence questionnaire aims at measuring a general experience of presence with or without other people. The factor analysis showed that there was a clear distinction between social presence and virtual presence so two specific phenomena are measured. The general conclusion is that communication medium does have a significant effect on the experience of virtual presence but that this effect might not be so large when interacting in a virtual environment as in a web environment.

The strongest conclusion in this thesis regarding the effects of communication medium on social presence is that people experience that voice and video/voice communication significantly increase the sensation of social presence compared to communicating with text-chat in a shared virtual environment. This means that they felt that the interaction was more social, personal and warm and that they felt more socially present, that they understood the other person's intentions, emotions and reactions and that they felt that their reactions were perceived by the other person. No significant differences were found when data from the voice and video/voice virtual reality conditions was compared to data from a voice web and a video/voice web condition. However, the pattern with an interaction effect was similar to the one for virtual presence, but the power was low for this analysis and did not reach significance. The general conclusion is that whether people use text-chat or a richer communication medium such as voice or video/voice makes a significant difference on perceived social presence in a shared virtual environment.

The conclusions regarding perceived performance in the shared virtual environment and different communication media are a bit confusing. Mean values showed that text-chat was most difficult, voice was a bit easier and that

video/voice was easiest to collaborate in but the only comparison that was significant was between text-chat and video/voice communication. The only other significant result was an interaction effect when data from the voice and video/voice virtual reality conditions was compared to data from a voice web and a video/voice web condition. The interaction effect showed that when participants interacted in a shared virtual environment, the task performance was perceived to be better in the video conference condition, but task performance was perceived to be better with audio in a web environment.

7.7 Methodological concerns and suggestions for further research

This thesis examines the use of modalities for collaboration in virtual environments in settings that are not commonly found in most people's everyday life. In certain specialized practices like flight simulations or training equipment for surgery these technologies have already been applied to a certain extent. Also, parts of the technology used in the thesis can be found in the computer games that children play with, like one degree of freedom haptics in joysticks and virtual environments that one navigates in, sometimes together with others. But overall, the research area is quite new and that is why the studies in the thesis have been experimental and performed in controlled settings. Therefore, great care has to be taken when generalizing the results to a real world situation. In future research the effects of different modalities in collaborative virtual environments need to be studied in real contexts where the collaboration has real consequences for users and where not only pairs of people are involved in the interaction but larger groups.

Only a very limited set of tasks have been studied in the thesis and future studies might reveal how important it is to match the modality used with the demands that each specific task puts on a person cognitively and physiologically. One question is how formal or informal the communication is in experimental situations. In one sense the communication is formal as people communicate on the initiative of a researcher and the participants are given a specific task to solve that constrains the topic of conversation. People usually have a preconception of what is expected of them in an experimental situation. On the other hand, if participants are randomly selected they do not belong to any specific organisation and they do not have any official rank or position in any organisational structure. The participants usually do not know each other beforehand and do not have a social history or relation to each other. The experimental situation might in fact appear to be quite accidental in the mind of the participants and the communication might not be perceived to be so formal after all, by the persons involved. It can be argued that one reason for

that is that it is unlikely that their contribution during the experiment, will have consequences for them personally in the future. But nevertheless, the dialogue is biased towards task oriented communication as the participants in the studies are instructed to solve a task.

The combinations of modalities and ways that these can be applied are endless and in the thesis only a small set of examples have been investigated. In the future people will also use these technologies together with their everyday artifacts and this will modify the design and functionality of the systems considerably. Yet, the conclusions drawn from these studies are useful for similar settings as the ones studied in the experiments.

Another concern has to do with the fact that quantitative measures have been used throughout the thesis. Only a few open interviews have been performed after the experimental sessions and my advice for future research is that interviews should always accompany these kinds of studies. My initial view was that it would be hard for people to talk about for example the touch modality as we are more used to discussing what we see and hear. It was especially hard to imagine that people would be able to express their opinion about the communicative aspects of haptic feedback. As soon as open interviews were made however, it became clear that people had quite a strong opinion about the benefits or disadvantages with different media and they expressed these in very clear words. One example is that users felt that they signaled to each other what they intended to do when they had haptic feedback but that this was not possible in the visual only interface. Suddenly the idea that haptic feedback can play a role in communication was felt to be more valid, and this assumption could then be further elaborated on. Such a conclusion would have felt far-fetched if it had only been based on the quantitative measures used in the thesis.

Questionnaires were used to a large extent in the thesis and there are a number of limitations and particularities that have to be considered with such an approach. The most urgent concern is the validity of the items that measure the concepts social presence, virtual presence and perceived performance. In this thesis the construct validity was tested using factor analysis in article C. It was found that the items in the social presence questionnaire measured a separate phenomenon than the two other questionnaires whereas the items measuring perceived performance overlapped with some of the items measuring virtual presence. The internal consistency measure of reliability was calculated for the three questionnaires, on data from the study presented in article C. The reliability was found to be good for all three questionnaires. Even though certain validity is gained by the theoretical background, interviews about the meaning of the concepts, or observations of behavior that is an indication of the phenomenon should be considered in future research. The dialogues between

the persons that collaborated during the experiments and their behaviors in the virtual environments could be analyzed in detail.

The scales used in the questionnaires in all studies were bipolar seven-point Likert type scales and they were later analyzed using ANOVA. In the studies ratings on all items in a questionnaire were summed for each person and the sum was then used in the analyses. Therefore mean values are not directly comparable for the three questionnaires, as the number of items in each questionnaire differs. But the mean values on a seven point scale are always reported in the article and they are of course comparable. There is a debate in natural science and social science about whether a seven point scale such as the one used can be analyzed using ANOVA. The established view in the area of social psychology is that ANOVA is a powerful and robust tool that can be used in order to analyze data like the questionnaire data obtained in the thesis and that it should be used.

Other statistical methods than the one used in the thesis could have been conducted in order to present a more comprehensive picture of users' collaboration and experiences of the interaction in the virtual environments. In the future the relation between actual task performance and perceived social presence, virtual presence and perceived performance could be further examined. Certain conclusions about causal relations can be drawn from the results in the studies, that for example haptic feedback generally improves precision and increases virtual and perceived performance. However, a correlational analysis comparing behavioral measures and people's ratings on questionnaires has not been performed and that would be an important relationship to analyze further in future research.

A few of the more intriguing results in the thesis are very tempting to follow up, as always. One result showed that the collaborative virtual environment makes a difference compared to a web environment even when rich media such as audio and video were available. Results showed that people felt a higher level of virtual presence when using video but only in the web conditions. Mean values showed that people using the collaborative virtual environment had a slow pace in the dialogue, had the most extensive conversation and spent a very long time interacting with each other. This is bewildering, the thesis cannot fully answer why people behave in this way but maybe future research in the area can clarify this issue.

Another open-ended result is the indirect indication that haptic feedback increased perceived virtual presence regardless if users could communicate with audio or not, whereas this was not true for perceived social presence. This shows again, that virtual presence and social presence are indeed two different phenomena and that it is important to measure them separately. It seems as if

haptic feedback is important for social presence and that certain communicative signals can be transmitted that way, but that audio again proves to be very important for social richness and therefore may overshadow the effect of haptic feedback. This is one of the most intriguing new hypotheses derived from the studies in the thesis and needs to be investigated more. This has also to do with the question of how much can be communicated haptically.

I hope that this thesis can show that it is possible to integrate old and novel technology in new ways and to utilize theories from research traditions in an interdisciplinary way in order to examine if for example shaking hands remotely makes a difference.

References

Allport, G. W. (1954). The historical background of modern social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology: Theory and method*, *Vol. 1* (pp. 1-80). Reading, MA: Addison-Wesley.

Anderson, A., Newlands, A., Mullen, J., Flemming, A. M., Doherty-Sneddon, G., & Van der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Journal of Interacting with Computers*, *8*, 193-206.

Appelle, S. (1991). Haptic perception of form: Activity and stimulus attributes. In M. Heller, & W. Schiff, (Eds.), *The psychology of touch*. (pp. 169-188). New Jersey: Lawrence Erlbaum Associates, Inc.

Argyle, M., & Dean, J. (1965). Eye contact, distance and affiliation. *Journal of Sociometry*, *28*, 289-304.

Axelsson, A-S. (2002). *The same as being there together? Communication, technology, and social context*. Licentiate Thesis. Department of Technology and Society, Chalmers University of Technology.

Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Cambridge, MA: Addison-Wesley.

Bannon, L., & Schmidt, K. (1989). CSCW: Four characters in search of a context. *Proceedings of the First European Conference on Computer Supported Cooperative Work* (pp. 358-372). Gatwick, UK.

Basdogan, C., Ho, C., Srinivasan, M. A., & Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction, 7*(4), 443-460.

Becker, B. & Mark, G. (2002). Social conventions in computer-mediated communication: A comparison of three online shared virtual environments. In R. Schroeder (Ed.), *The Social Life of Avatars* (pp. 19-39). London: Springer.

Biocca, F., Burgoon, J. Harms, C., & Stoner, M. (2001). Criteria and scope conditions for a theory and measure of social presence. *Proceedings of the 4th Annual International Workshop Presence'01*. Available online (2004-02-25) at http://astro.temple.edu/~lombard/P2001/papers.htm

Biggs, S. J., & Srinivasan, M. A. (2002). Haptic interfaces. In K. M. Stanney (Ed.), *Handbook of virtual environment technology*. (pp. 93-116). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Blascovich, (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars* (pp. 127-145). London: Springer.

Bly, S. A., Harrison, S. R., & Irwin, S. (1993). Media spaces: Video, audio and computing. *Communications of the ACM*, *36*, 1, 28-47.

Bowers, J., O'Brien, J., & Pycock, J. (1996). Practically accomplishing immersion: Cooperation in and for collaborative virtual environments. In M. S. Ackerman (Ed.), *Proceedings of CSCW'96* (pp. 380-389). New York: ACM Press.

Brave, S., Ishii, H., & Dahley, A. (1998). Tangible interfaces for remote collaboration and communication. *Proceedings of CSCW'98* (pp. 169-178). New York: ACM Press.

Burdea, G. C. (1996). *Force and touch feedback for virtual reality*. New York: Wiley.

Burrell, G., & Morgan, G. (1993). *Sociological paradigms and organisational analysis.* Brookfield, VT: Ashgate Publishing.

Card, S.K., English, W.K., & Burr, B.J. (1978). Evaluation of mouse, rate controlled isometric joystick, step keys and text keys for text selection on a CRT. *Journal of Ergonomics*, *21*, 601-613.

Chang, A., O'Modhrain, S., Jacob, R., Gunther, E., & Ishii, H. (2002). ComTouch: Design of a vibrotactile communication device. *Proceedings of the Conference on Designing Interactive Systems: Processes, practices, methods, and techniques, DIS'02* (pp. 312-320). London, UK.

Chapanis, A. (1975). Interactive human communication. *Journal of Scientific American*, 232, 36-42.

Chapanis, A. (1981). Interactive human communication: Some lessons learned from laboratory experiments. In B. Shackel, (Ed.), *Man-computer interaction: Human factors aspects of computers and people*. (pp. 65-114). Alphen aan den Rijn: Sijthoff & Noordhoff

Cherry, C. (1957). *On human communication*. Cambridge: MIT Press.

Clark, H. H., & Brennan, S. E., (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared Cognition* (pp. 127-149). Washington: American Psychological Association.

Clark H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Journal of Cognition*, 22, 1-39.

Daft, R. L. (1998). *Essentials of organization theory and design*. Cincinnati, OH: South-Western College Publishing.

Daft, R. L., & Lengel, R. (1984). Information richness: A new approach to managerial behavior and organization design. In L. L. Cummings, & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 6, pp. 191-233). Greenwich, CT: JAI

Daft, R. L., & Lengel, R. (1986). Organizational information requirements, media richness, and structural design. *Journal of Management Science*, 32, 554-571.

Daft, R. L., Lengel, R., & Trevino, L. K. (1987). Message equivocality, media selection and manager performance: Implications for information systems. *Journal of MIS Quarterly*, 11, 355-366.

Daly-Jones, O., Monk, A., & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: Fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*, 49, 21-58.

Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. In P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'92* (pp. 541-547). New York: Association for Computing Machinery ; Reading, MA: Addison-Wesley.

Durlach, N., & Slater, M. (2000). Presence in shared virtual environments and virtual togetherness. *Journal of Presence: Teleoperators and Virtual Environments*, 9(2), 214-217.

Egido, C. (1990). Teleconferencing as a technology to support cooperative work. In J. Galegher, R. Kraut, & C. Egido (Eds), *Intellectual teamwork* (pp. 351-371). Hillsdale, NJ: Lawrence Erlbaum.

El-Shinnawy, M., & Markus, M. L. (1997). The poverty of media richness theory: Explaining people´s choice of electronic mail vs. voice mail. *International Journal of Human-Computer Studies*, 46, 443-467.

Exline, R. V., Gray, D., & Schuette, D. (1965). Visual behavior in a dyad as affected by interview content and sex of resondent. *Journal of Personality and Social Psychology*, *1*, 201-209.

Finn, K., Sellen, A., & Wilbur, S. (1997). *Video mediated communication*. Hillsdale, NJ: Lawrence Erlbaum Associated.

Fish, R. S., Kraut, R. E., & Chalfonte, B. L. (1990). The video window system in informal communication. *Proceedings of CSCW'90* (pp.). New York: The Association for Computing Machinery.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381-391.

Fogg, B. J., Cutler, L., Arnold, P., & Eisback, C. (1998). HandJive: a device for interpersonal haptic entertainment. In C-M. Karat & S. Pemberton (Eds.), *Proceedings of CHI'98* (pp. 57-64). Reading: Addison-Wesley.

Fontaine, G. (1992). The experience of a sense of presence in intercultural and international encounters. *Journal of Presence: Teleoperators and Virtual Environments, 1*(4), 482-490.

Freeman, J., Avons, S.E., Harrison, D., Lodge, N., & Pearson, D., (1998). *Behavioural realism as a metric of presence*. Available online (2004-02-24) at http://homepages.gold.ac.uk/immediate/btpres/BTpres.htm

Freides, D. (1974). Human information processing and sensory modality: Crossmodal functions, information, complexity, memory and deficit. *Psychological Bulletin*, *81*, 284-310.

Galbraith, J. (1973). *Designing complex organizations*. Reading, MA: Addison-Wesley.

Gaver, W. W., Smith, R. B., & O´Shea, T. (1991). Effective sounds in complex systems: The ARKola simulation. In S. P. Robertson, G. M. Olson & J. S. Olson (Eds.), *Proceedings of CHI´91* (pp. 361-367), New York: Association for Computing Machinery.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Goffman, E. (1963). *Behavior in public places: Notes on the social organization of gatherings*. London: Free Press of Glencoe.

Gupta, R., Sheridan, T., & Whitney, D. (1997). Experiments using multimodal virtual environments in design for assembly analysis. *Journal of Presence: Teleoperators and Virtual Environments, 6* (3), 318-338.

Hage, J., Aiken, M., & Marrett, C. (1971). Organization structure and communication. *Journal of American Sociological Review*, *36*, 860-871.

Hannaford, B., Wood, L., McAffee, D., & Zak, H. (1991). Performance evaluation of a six-axis generalized force-reflecting teleoperator. *IEEE Transactions on Systems, Man and Cybernetics, 21*(3), 620-633.

Hasser, C. J. Goldenberg, A. S., Martin, K. M., & Rosenberg, L. B. (1998). User performing a GUI pointing task with a low-cost force-feedback computer mouse. *Proceedings of the ASME Dynamic Systems and Control Division* (Vol. 64, pp. 151-156).

Heeter, C. (1992). Being there: The subjective experience of presence. *Journal of Presence: Teleoperators and Virtual Environments*, *1*(2), 262-271.

Heilbronn, M., & Libby, W. L. (1973). *Comparative effects of technological and social immediacy upon performance and perception during a two-person game*. Paper presented at the Annual Convention of the American Psychological Association, Montreal.

Heller, M., & Schiff, W. (1991). *The psychology of touch*. New Jersey: Lawrence Erlbaum Associates, Inc.

Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Journal of Presence: Teleoperators and Virtual Environments, 5*(3), 290-301.

Hubbold, R. J. (2002). Collaborative stretcher carrying: A case study. *The 8th Eurographics Workshop on Virtual Environments* (pp. 7-12).

Hurmuzlu, Y., Ephanov, A., & Stoianovici, D. (1998). Effect of a pneumatically driven haptic interface on the perceptional capabilities of human operators. *Journal of Presence: Teleoperators and Virtual Environments*, 7(3), 290-307.

Ijsselsteijn, & Ridder, (1998). Measuring temporal variations in presence. *Proceedings of the Presence in Shared Virtual Environments Workshop, First International Workshop on Presence*. Ipswich, Suffolk, UK.

Ishii, M., Nakata, M., & Sato, M. (1994). Networked SPIDAR: A networked virtual environment with visual, auditory and haptic interactions. *Journal of Presence: Teleoperators and Virtual Environments*, 3(4), 351-359.

ISO, ISO 9241-9, *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices (FDIS - Final Draft International Standard),* 2000, International Organization for Standardization.

Jensen, C., Farnham, S. D., Drucker, S. M., & Kollock, P. (2000). The effect of communication modality on cooperation in online environments. In T. Turner et al. (Eds.), *Proceedings of the ACM Conference on Computer-Human Interaction 2000* (pp. 470-477). New York : ACM Press.

Jordan, J., Mortensen, J., Oliveira, M., Slater, M., & Srinivasan, M. (2002). Collaboration in a mediated haptic environment. *The 5th Annual International Workshop on Presence*. Porto, Portugal.

Jää-Aro, K-M. (2004). *Reconsidering the avatar: From user mirror to interaction locus*. Doctoral Thesis, TRITA NA-0403. Royal Institute of Technology, Stockholm, Sweden.

Katz, D. (1989). *The world of touch*. (L. E. Krueger, Trans.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Katz, R., & Tushman, M. (1978). Communication patterns, project performance, and task characteristics: An empirical evaluation in an R & D setting. *Journal of Organizational Behavior and Human Performance*, 23, 139-162.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39, 1123-1134.

Kiesler, S., Zubrow, D., Moses, A., & Geller, V. (1985). Affect in computer-mediated communication: An experiment in synchronous terminal-to-terminal discussion. *Journal of Human-Computer Interaction*, 1, 77-104.

Kraut, R. E., Fish, R. S., Root, R. W., & Chalfonte, B. L., (1993). Informal communication in organizations: Form, function and technology. In R. M. Baecker, (Ed.), *Readings in groupware and computer-supported cooperative work: Assisting human-human collaboration* (pp. 287-314). San Mateo, CA: Kaufmann.

Kraut, R., Galegher, J., Fish, R., & Chalfonte, B. (1992). Task requirements and media choice in collaborative writing. *Journal of Human-Computer Interaction*, *7*, 375-407.

Lederman, S.J., & Klatzky, R.L. (1987). Hand movements: A window into haptic object recognition. *Journal of Cognitive Psychology*, *19*, 342-368.

Lawrence, P., & Lorsch, J. (1967). *Organizations and environment*. Cambrige, MA: Harvard University Press.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2). Available online (2004-02-24) at http://www.ascusc.org/jcmc/vol3/issue2/lombard.html

Loomis, J. M., & Lederman, S. J. (1986). Tactual perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Cognitive processes and performance* (Vol. 2). New York: Wiley.

Malone, T.W., & Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems. In F. Halasz (Ed.), *Proceedings of the Conference on Computer Supported Cooperative Work, CSCW'90*. (pp. 357-370). Los Angeles, CA: ACM.

Mason, A. H., Walji, M. A., Lee, E. J., & MacKenzie, C. L. (2001). Reaching movements to augmented and graphic objects in virtual environments. *Proceedings of Conference on Human Factors in Computing Systems, CHI'01* (pp. 426-433). New York: ACM Press.

Matsuura, N., Fujino, G., Okada, K., & Matsushita, Y. (1993). An approach to encounters and interaction in a virtual environment. *Proceedings of ACM Computer Science Conference* (pp. 298-303).

McGrath, J. E. (1984). *Groups: Interaction and performance*. Engelwood Cliffs, NJ: Prentice-Hall.

McGrath, J. E. (1993). Time, interaction and performance (TIP): A theory of groups. Small group research. In R. M. Baecker (Ed.), *Readings in groupware and computer-supported cooperative work : assisting human-human collaboration* (pp. 116-129). San Mateo, CA: Kaufmann.

McGrath, J. E., & Hollingshead, A. B. (1994). *Groups interacting with technology*. Thousand Oaks: SAGE Publications.

MacKenzie, I.S. (1992). Fitts' law as a research and design tool in human computer interaction. *Journal of Human Computer Interaction*, *7*, 91-139.

MacKenzie, I.S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. In S. P. Robertson, G. M. Olson & J. S. Olson (Eds.), *Proceedings of the Conference on Human Factors in Computing Systems, CHI'91* (pp. 161-166). New York: Association for Computing Machinery.

McLaughlin, M., Sukhatme, G. S., Peng, W., Zhu, W., & Parks, J. (2003). Performance and co-presence in heterogeneous haptic collaboration. *The eleventh Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems, IEEE Virtual Reality'03*.

McLeod, P. L. (1996). New communication technologies for group decision making: Toward an integrated framework. In R. Y. Hirokawa, & M. S. Poole (Eds.), *Communication and group decision making*. (pp.426-461).Thousand Oaks: Sage Publications, Inc.

Mehrabian, A. (1969). Methods and design: Some referents and measures of nonverbal behavior. *Journal of Behav. Res. Meth. and Instrum.*, *1*(6), 203-207.

Mehrabian, A. (1972). *Nonverbal Communication*. Chicago: Aldine.

Minsky, M. (1980). Telepresence. *Omni*. NY: Omni Publications International Ltd.

Oakley, I., Brewster, S., & Gray, P. (2001). Can you feel the force? An investigation of haptic collaboration in shared editors. In C. Baber, M. Faint, S. Wall & A. M. Wing (Eds.), *Proceedings of Eurohaptics 2001* (pp. 54-59). Birmingham.

Oakley, I., McGee, M. R., Brewster, S., & Gray, P. (2000). Putting the feel in "look and feel". *ACM CHI Letters,* 2(1), 415-422.

Ochsman, R.B., & Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during co-operative problem-solving. *International Journal of Man-Machine Studies*, *6*, 579-619.

O'Connaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Journal of Human Computer Interaction*. *8*, 389-428.

Olson, J. S., Olson, G. M., & Meader, D. K. (1995). What mix of video and audio is useful for small groups doing remote real-time design work? In I. R. Katz, R. Mack, & L. Marks (Eds.), *Proceedings of the Conference on Human Factors in Computing Systems, CHI´95* (pp. 362–368). New York: ACM Press.

O'Modhrain, S. (2001). *Playing by feel: Incorporating haptic feedback into computer-based musical instruments*. Doctor of Philosophy Thesis. Stanford University.

Poole, M. S., Seibold, D. R., & McPhee, R. D. (1996). The structuration of group decisions. In R. Y. Hirokawa, & M. S. Poole (Eds.), *Communication and group decision making*. (pp.114-146). Thousand Oaks: Sage Publications, Inc.

Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1085-1099.

Rice, R. E. (1993). Media appropriateness: Using social presence theory to compare traditional and new organizational media. *Journal of Human Communication Research*, *19*(4), 451-484.

Ruddle, R. A., Savage, J.C.D., & Jones, D.M. (2002). Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction*.

Rutter, D., & Stephenson, G. (1977). The role of visual communication on synchronising conversation. *European Journal of Social Psychology*, *2*, 29-37.

Salisbury, K., Brock, D., Massie, T., Swarup, N., & Zilles, C., (1995). Haptic rendering: Programming touch interaction with virtual objects. *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (pp. 123-130). New York: ACM Press.

Sallnäs, E-L. (submitted). Passing virtual objects collaboratively with and without haptic feedback: Effects on social presence, virtual presence and perceived performance.

Sallnäs, E-L. (conditionally accepted). Effects of communication mode on social presence, presence and performance in collaborative virtual environments. *Journal of Presence: Teleoperators and Virtual Environments*.

Sallnäs, E-L. (2001). Improved precision in mediated collaborative manipulation of objects by haptic force feedback. In G. Goos, J. Hartmanis and J. van Leeuwen (Series Eds.) and S. Brewster and R. Murray-Smith (Vol. Eds.),

Lecture Notes in Computer Science: Vol. 2058. Haptic Human-Computer Interaction (pp. 69-75). Heidelberg, Germany: Springer.

Sallnäs, E-L., Rassmus-Gröhn, K., & Sjöström, C. (2000). Supporting presence in collaborative environments by haptic force feedback. *ACM Transactions on Computer-Human Interaction*, *7*(4), 461-476.

Sallnäs, E-L., & Zhai, S. (2003). Collaboration meets Fitts' law: Passing virtual objects with and without haptic force feedback. In M. Rauterberg, M. Menozzi & J. Wesson (Eds.), *Proceedings of INTERACT´03* (pp. 97-104). Amsterdam: IOS Press.

Sauter, C., Morger, O., Mühlherr, M., Hutchison, A., & Teufel, S. (1995). CSCW for strategic management in Swiss enterprizes: an empirical study. In H. Marmolin, Y. Sundblad, & K. Schmidt (Eds.), *Proceedings of the Fourth European Conference on Computer Supported Cooperative Work* (pp.117-132.). Netherlands: Kluwer Academic Publishers.

Schein, E. H. (1965). *Organizational psychology*. New Jersey: Prentice-Hall, Inc.

Schloerb, D. W. (1995). A quantitative measure of telepresence. *Journal of Presence: Teleoperators and Virtual Environments, 4*(1), 64-80.

Schroeder, R. (2002). Social interaction in virtual environments: Key issues, common themes, and a framework for research. In R. Schroeder (Ed.), *The social life of avatars* (pp. 1-18). London: Springer.

Sellen, A. J. (1992). Speech patterns in video-mediated conversations. In P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'92* (pp. 49-59). New York: Association for Computing Machinery.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. London: Wiley.

Simmel, G. (1921). Sociology of the senses: Visual interaction. In. R.E. Park, & E. W. Burgess (Eds.), *Introduction to the science of sociology* Chicago: University of Chicago Press.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small group behaviour in a virtual and real environment: A comparative study. *Journal of Presence: Teleoperators and Virtual Environments, 9*(1), 37-51.

Slater, M., & Steed, A. (2002). Meeting people virtually: Experiments in shared virtual environments. In R. Schroeder (Ed.), *The social life of avatars* (pp. 146-171). London: Springer.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Journal of Presence: Teleoperators and Virtual Environments, 6*(6), 603-616.

Smith, R. L., Richetto, G. M., & Zima, J. P. (1972). Organizational behaviour: An approach to human communication. In R. W. Budd (Ed.), *Approaches to human communication* (pp. 269-289). New York: Spartan Books.

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Journal of Management Science*, *32*, 1492-1512.

Stephenson, G. M., Ayling, K., & Rutter, D. R. (1976). The role of visual communication in social exchange. *British Journal of Social and Clinical Psychology*, *15*, 113-120.

Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Journal of Human-computer Interaction*, *12*(3), 227-266.

Strong, R., & Gaver, B. (1996). Feather, scent and shaker: Supporting simple intimacy. In M. S. Ackerman (Ed.), *In Videos, Demonstrations and Short Papers of CSCW'96* (pp. 29-30). New York: Association for Computing Machinery.

Tang, J. C., & Isaacs, E. (1993). Why do users like video? Studies of multimedia supported collaboration. *Journal of Computer Supported Cooperative Work*, *1*, 163-196.

Van de Ven, A.H., Delbecq, A.L., & Koenig, R.Jr. (1976). Determinants of coodination modes within organizations. *American Sociological Review*, *41*, 322-338.

Vaske, JJ., & Grantham, C.E., (1989). *Socializing the human-computer environment*. Norwood: Ablex.

Veinott, E. S., Olson, J., Olson, G. M., & Fu, X. (1999). Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. In M. G. Williams et al. (Eds.), *Proceedings of the Conference on Human Factors in Computing Systems, CHI'99* (pp. 302-309). New York : ACM.

Wang, Y., & MacKenzie, C.L. (2000). The role of contextual haptic and visual constraints on object manipulation in virtual environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'00* (pp. 532-539). New York, NY: ACM Press.

Williams, W. (1977). Experimental comparisons of face-to-face and video mediated communication. *Psychological Bulletin*, *84*, 963-976.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Journal of Presence: Teleoperators and Virtual Environments*, *7*(3), 225-240.

Whittaker, S., & O-Conaill, B. (1997). The role of vision in face-to-face and mediated communication. In E. E. Finn, A.J. Sellen, & S.B. Wilbur (Eds.), *Video-Mediated Communication*. (pp. 23-50). Mahwah, NJ: Lawrence Erlbaum.

Zhai, S. (2002). *On the validity of throughput as a characteristic of computer input*, IBM Research Report RJ 10253, Almaden Research Center: San Jose, California.