

# Auditory Feedback in Haptic Collaborative Interfaces

YING YING HUANG, JONAS MOLL, EVA-LOTTA SALLNÄS, YNGVE SUNDBLAD

HCI Dept., School of Computer Science and Communication, Royal Institute of Technology, Sweden

The combined effect of haptic and auditory feedback in shared interfaces, on the cooperation between visually impaired and sighted persons is under-investigated. A central challenge for cooperating group members lies in obtaining a common understanding of the elements of the workspace and maintaining awareness of the other members', as well as one's own, actions during the work process. The aim of the experimental study presented here, was to investigate if adding audio cues in a haptic and visual interface makes collaboration between a sighted and a blindfolded person more efficient. Results showed that task performance was significantly faster in the audio, haptic and visual feedback condition compared to the haptic and visual feedback condition. One special focus was also to study how participants utilize the auditory and haptic force feedback in order to obtain a common understanding of the workspace and to maintain an awareness of the group members' actions. Results from a qualitative analysis showed that the auditory and haptic feedback was used in a number of important ways for the group members' action awareness and in the participants' grounding process.

## **Categories and Subject Descriptors:**

H.5.2 [Information Interfaces and Presentation]: User Interfaces-Auditory (non-speech) feedback, Haptic I/O , Evaluation/methodology; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces-Evaluation/methodology; Synchronous interaction

**General Terms:** Experimentation, Human Factors, Measurement, Performance

## **Additional Key Words and Phrases:**

Haptic, Force feedback, Virtual environments, Multimodal interface, Collaboration, Awareness, Common ground

Corresponding author: Ying Ying Huang

E-mail address: [yingying@csc.kth.se](mailto:yingying@csc.kth.se)

- 
- 

## 1 INTRODUCTION

During computer-supported cooperation, a central challenge lies in increasing peoples' attention to actions done by those involved in joint work in a shared interface. It is particularly interesting to investigate how touch and audio feedback in combination support communication and interaction between people during collaboration in shared interfaces in a situation where one person cannot see. Computer supported group work is a much neglected perspective in assistive research. A haptic device that provides force feedback can assist users in getting feedback to the

touch modality and in identifying objects in a computer interface. Here, haptic feedback refers to an integration of both kinesthetic sensing (i.e. of the position and movement of joints and limbs) and tactile sensing (i.e. through the skin) [Loomis and Lederman 1986]. Auditory feedback refers to the use of non-verbal sound to convey information to users in a computer interface.

In an earlier study [Sallnäs et al. 2007] a collaborative application, aiming at supporting visually impaired pupils in learning about spatial geometry in group work with sighted peers in primary school, was developed and evaluated. In that application pupils could feel the shape of different geometrical objects, which could be picked up and moved around using a touch feedback system. The pupils could build composed objects out of smaller ones. The application was evaluated in groups of two sighted children and one visually impaired child in four elementary schools in the Stockholm area. The results shed light on several important aspects of haptic feedback when it came to support for collaboration between sighted and visually impaired children. The need for reference points was highlighted and it was shown that the haptic feedback had a positive effect on the inclusion of the visually impaired pupil. Haptic guiding functions, by which the sighted pupil could physically drag the visually impaired pupil around, were also shown to be a great aid. However, the results from the evaluation showed that it was problematic to maintain action awareness in this kind of dynamic environment where all users could move objects around. No audio cues were heard that indicated that a person interacted with an object.

Based on the findings from that previous evaluation, four different audio cues were added to the interface. The resulting application was evaluated in the study presented in this paper with pairs of sighted university students of which one was blindfolded. The reason for using sighted students was to get enough subjects to be able to test the hypothesis that audio cues make collaboration faster because of improved action awareness in a haptic collaborative virtual environment. In one of the two conditions, participants could use auditory feedback that gave information about actions performed in the virtual collaborative environment. In both conditions (with or without auditory feedback), participants sat beside each other in the same room and could communicate verbally and touch and manipulate objects in a visual interface with one haptic device each.

Several questions regarding users' action awareness were considered in the present study. How do sighted and non-sighted participants attend to the other person's actions during collaborative work in a virtual environment? Can the level of action awareness participants get affect the grounding strategies they use?

## 2 THEORETICAL BACKGROUND

### 2.1 Haptic Force Feedback

Haptic technologies have become widely available and inexpensive, and the advantage of being able to use the touch modality is being increasingly recognized. The research concerning haptic perception and rendering techniques has increased rapidly during the last few years, and results have shown the significant role that haptic feedback plays in graphical single user interfaces [Gupta et al. 1997; Hasser et al. 1998; Hurmuzlu et al. 1998]. Users with a severe visual impairment have to work without the visual modality, which restrains their ability to utilize graphical user interfaces. Specifically, it is harder using the sense of touch to get an overview of an environment and localize and explore objects and interesting parts thereof [Jansson 2007].

Designers are beginning to realize the advantages of haptic displays when it comes to helping blind individuals to overcome the challenges experienced when accessing and exploring the web [Kuber 2007]. The touch modality has also been shown to make it possible for visually impaired to explore and navigate virtual environments [Sjostrom 2001]. The interaction is enriched by the use of the sense of touch, since visually impaired users can identify objects and perceive their shape and texture. In a recent EU project, Pure Form, the aim was to develop a haptic display for exploration of virtual copies of statues at museums in order to make them accessible to visually impaired people, [Bergamasco et al. 2001; Bergamasco and Prisco 1998; Frisoli et al. 2002].

In a number of studies on collaboration between sighted people, it has been shown that haptic feedback improves task performance and increases perceived presence and the subjective sense of togetherness for different application areas in shared virtual environments [Ho et al. 1998; Basdogan et al. 1998; Durlach and Slater 2000; Sallnäs et al. 2000; Basdogan et al. 2000; Oakley et al., 2001]. Less attention has been paid to the impact of auditory feedback in combination with haptic feedback in a collaborative setting with sighted and persons that cannot see. However, in one study by Plimmer the interaction between visually impaired pupils and their teachers was addressed [Plimmer et al. 2008]. This study investigated the effects of training handwriting using haptic and audio output to realize a teacher's pen input to the pupil. Results showed that training in the system improved the character shapes drawn by visually impaired, especially those that were completely blind.

## 2.2 Auditory Feedback

Audio displays are also getting increased attention for their capability to support navigation and collaboration in virtual environments.

A number of researchers have shown how 3D virtual environments can be made accessible to sighted and visually impaired people by providing auditory feedback. Winberg and Hellström [2001] developed an interface based solely on auditory feedback. They developed a sound model that made it possible for blind users to play the game "Towers of Hanoi". The game was set with either three or four disks and each disk had its unique sound differing in pitch and timbre from the others. The height on the peg of a particular disk was represented by the length of the sound. Stereo panning was used to convey information about which peg a particular disk was on. The results showed the potential of the audio modality to convey information to visually impaired users, who could play the game together with a sighted person in a way that included both players in the process of solving the problem.

In an experiment conducted by Poll and Eggen [1996], a blind subject used an absolute mouse to scan for graphical user interface (GUI) objects represented by speech and non-speech sounds within a rectangular area bounded by standing edges. Kennel [1996] suggested that blind users, within a relatively short time could read simple diagrams with the aid of a touch panel and an associated auditory display. The diagram was displayed on a sheet of paper covering a tablet. When touched, parts of it generated relevant audio messages. The whole diagram could thus be explored using this audio-tactile strategy.

According to the design principles of Winberg [2004], a non-visual interface should provide access to all functionality. In principle all the things that can be done by a sighted user in a GUI should be possible to do with an interface for a blind user (this will include giving a sense to the

'iconicity' of the interface and its spatial arrangement). Furthermore, it should enable manipulation and exploration. If possible the direct manipulation methods of GUIs should be given a parallel implementation. Finally, it should be coherent with visual interfaces in order to enable collaboration between sighted and visually impaired users. This is likely to involve the cross-media translation of significant interface elements for creating a common mapping of the elements used in both kinds of interfaces.

An approach of transforming visual information into non-visual media has been used for mapping GUI objects in a study by Crommentuijn [2006]. Five different ways of representing a set of objects for visually impaired users in a GUI were implemented in an auditory and haptic interface and visually impaired users' interaction with this interface was investigated. It was shown that the design in which the user could "hold a virtual microphone" and move around until objects were found was the most efficient one.

From the results in these studies it can be implied that information such as the location of objects and the location and action of your partner in a collaborative visual or haptic context could also be represented and conveyed by auditory cues. However, less attention has been paid to the impact of auditory feedback in combination with haptic feedback in a collaborative setting with sighted and persons that cannot see.

### 2.3 Assistive Technology for Collaboration

In Swedish schools pupils often do group work in many different study subjects. This pedagogical approach trains social skills and supports collaborative learning. Assistance in this group work process is particularly important if one of the pupils is visually impaired since the most important sense – vision – is not available.

There are several studies that have investigated issues regarding collaboration between visually impaired people in educational settings. McGookin and Brewster [2007] investigated collaboration between visually impaired users during interactive browsing and manipulation of simple graphs. They specifically looked at supporting awareness of others' activities [McGookin and Brewster 2007].

The teaching and learning of subjects in small groups, largely without direct teacher involvement, has been argued to improve the social, academic and cognitive abilities of students [Slavin and Cooper 1999]. For visually impaired students, however, there are significant problems in accomplishing these aims. In interviews with visually impaired students, Sallnäs et al. [2007] describe how students enjoyed group work, because it allowed students to "get to know other pupils better". Sallnäs et al. [2007] also found however, that it was difficult for students to keep track of others' activities. Because the sighted and visually impaired students used different representations of information, it was difficult for example to point out something to the other person in the same way two sighted users would. This makes it difficult for each person to know what his/her partner is looking at, or to be sure that both are referring to the same information.

### 2.4 Awareness in Dynamic 3D Environments

In real life we can have an extensive overview of the environment we are in and of objects in it. However, we always focus on some objects which attract our attention. Moreover, we attend to objects by multiple senses. In virtual environments, multisensory experiences are created through

artificial means and the effectiveness of virtual environments has often been linked to the richness of the sensory information provided and the realness of the experience [Basdogan et al. 2000; Held and Durlach 1992; Ellis 1992; Barfield and Furness 1995].

The extent to which users are aware of things that happen in virtual environments is of great importance. Awareness is generally used in terms of individuals' perception of others' activities and the status of others' work processes [Dourish and Bellotti 1992]. A key factor to consider in collaborative virtual environments is the way in which continuous awareness of others' activities allows people to manage their own activity in social situations in a flexible way and to predict actions of others [Benford et al. 1994]. When people cooperate without having the opportunity to get sufficient awareness information, they do not reach the same quality in joint projects [Kraut et al. 1993].

In order to clarify terminology on awareness, we refer to Carroll et al. [2003], who distinguish between three kinds of awareness in virtual settings and emphasize that each type can be supported by certain tools. Social awareness refers to the users' consciousness of the presence of others. Social awareness can be fostered by tools which visualize the presence of others in any way, for example by using photographs of the team members. Tools fostering action awareness provide information about the actions which are currently being carried out by the group and its members, for example by showing which shared resources (e.g. a document) other team members are interacting with. Accordingly, action awareness is primarily important in the context of synchronous collaboration. Activity awareness focuses on the task that is to be performed by the group. Here, the actions of the participants involved are related to the mutual task, but activity awareness is primarily important in the context of asynchronous collaboration. Carroll et al. [2003] presented this theory of awareness in the context of computermediated communication in a distributed environment. Many studies have paid attention to awareness and its impact in distributed virtual reality systems [Benford et al. 1994; Kraut 1993; Kimmerle 2007] and in the context of computer supported collaborative work [Schmidt 2002; Neale et al. 2004; Carroll et al. 2006].

We believe that the concept of action awareness is also relevant in a co-located situation, especially when one person cannot see. When users are working on a joint task in a shared virtual environment, such as for example putting together a model of a machine in a shared interface, people not only need to communicate verbally. They need to be aware of the location of the others and of the actions they take and especially their interaction with objects in the shared workspace. In a collocated situation, action awareness of the things that group participants do in the physical as well as the virtual environment has to be attended to by all participants in order to be able to perform the task.

In collaborative situations, sighted people mostly rely on visual information whereas visually impaired people mainly maintain awareness by touching and hearing. It is necessary to transfer visual information, at least to some extent, into non-visual representations in order to make computer-supported communication and collaboration between sighted and visually impaired people possible. In this case getting action awareness depends on multimodal technical support. Haptic and auditory feedback systems give sensory information that potentially can provide crucial cues for visually impaired users that make mutual awareness during collaborative work possible to achieve.

## 2.5 Shared Understanding of the Work Space

Clark and Brennan [1991] define common ground as a state of mutual understanding among conversational participants about the topic at hand.

Although common ground is a general theory of language use, it holds true for all collaborative activities that people must update their common ground on a continuous basis, and do so through a grounding process [Neale et al. 2004]. To communicate, collaborate, and coordinate, people must share a vast amount of information and mutual knowledge. We argue that group members can achieve common ground in a variety of ways, not only through conversation but also by different sensory information like touch and interaction sounds.

Collaboration depends on an unrestrained dialogue and that people working together can maintain a common ground of the context of joint action including the features of the workspace and the continuous status of the work process [Clark 1996]. Some of the basic features of face-to-face conversation are that people share a physical environment, that they have the ability to see and hear each other and that they can perceive each other's actions. In the case of group work between visually impaired and sighted pupils in school today, a shared understanding cannot always be maintained of either the workspace or the work process. This is due to the fact that their respective work tools do not allow mutual access to the same information or perception of others' actions in it [Sallnäs et al. 2007].

Common ground is established and maintained through dialogues between collaborators and grounding activities aim to provide mechanisms that enable people to establish and maintain common ground [McCarthy et al. 1991]. An example is the use of deictic references [Clark 2003; Clark 1996] to i.e. an object, place or person. When establishing common ground the use of deixis (references to specific objects) is important. Deictical references, like "that", "this", "the big one" are used to direct the partner's attention to a specific object or part of the interface. Maintaining common ground is also shown to be much easier when collaborators can make use of this kind of references [Burke and Murphy 2007]. The ability to use gestures has been shown to be a considerable aid in the grounding process [Kirk et al. 2007]. If one of the collaborating partners is visually impaired, visual gestures cannot be used. We will in this article argue, that audio cues could convey grounding information and can be communicative similarly to gestures. Gestures can be used to create meaningful information to others [Clark, 2003]. The importance of providing the possibility for deictic referencing in collaborative environments has been acknowledged in a number of studies [Cherubini 2008; Ou et al. 2003; Kirk et al. 2007; Fussell et al. 2004]. Apart from pointing Clark [2003] argues that placing oneself in relation to other persons and the context as well as placing objects are equally essential indicative acts as pointing. The implication of interaction sound for communication is clear if it for example is possible to hear when an object is put down or picked up.

In the current study, the central questions are how multiple sensory cues with haptic and auditory grounding tools can improve mutual understanding in order to enrich communication and enhance coordination, and how they potentially make collaboration more efficient in a situation where one of the persons cannot see the workspace. To our knowledge, the impact of audio feedback on efficiency and satisfaction in haptic collaborative interfaces has not been sufficiently investigated thus far.

### 3 AIM AND HYPOTHESES

In our research we want to investigate whether adding audio functions in a visual and haptic interface will make a difference beyond the impact of haptic feedback. The hypothesis tested in this experiment was that adding audio functions to a collaborative visual/haptic interface would improve task performance in a collaborative haptic 3D virtual environment. We argue that improved action awareness will make collaboration faster and will facilitate the grounding process. Examples of this are shown in the qualitative analysis in this paper.

(H1) Adding audio feedback into the collaborative visual/haptic environment will make task performance faster.

### 4 METHOD

#### 4.1 Experimental Design

A between subjects design was used in this experiment, with two conditions: (1) a visual and haptic VR environment, and (2) an audio, visual and haptic VR environment.

The dependent variable was task performance. Task performance was measured by the time spent by group members to solve a task during the test.

The test sessions ended with an open form of interview with each pair. Questions were asked in the interview about the subjects' perception of the system, with special focus on awareness, common ground, and joint task performance in different modalities. An observation analysis of the video recordings from the experimental sessions was also made in order to get a more detailed understanding of how audio cues affected the interaction.

#### 4.2 Subjects

Twenty-eight university students participated in this experiment and they were collaborating in 14 pairs. All subjects were sighted but one subject was blindfolded in each pair when the test task was performed. Seven of the pairs used the interface with audio, visual and haptic force feedback and the other seven pairs used the interface with only visual and haptic force feedback.

Visually impaired persons were not recruited to this experiment even though it would have been better than blindfolding sighted people. More participants were needed, than could be recruited among visually impaired persons. Results from an earlier study [Sallnäs et al. 2007] have shown how visually impaired pupils managed to collaborate with sighted peers in a very similar version of the haptic environment. In the study presented here only one single variable was added namely the audio cues. As the haptic and visual application was found to work well for visually impaired persons, except for the problem with awareness of changes made in the environment, we argue that the effects found in this experimental study will be possible to generalize to a large extent to visually impaired persons. The general level may be different, but if a parameter has an effect on sighted people, it can be expected to also have an effect on visually impaired people.

### 4.3 The Application

The collaborative interface used in this experiment was modeled to be perceived as a room viewed from above through a transparent ceiling (Figure 1). The collaborative setting of the experiment is shown in Figure 2. The room contained cubes that could be picked up and moved around by means of touch feedback using a haptic feedback device called Phantom. The roof, walls and the cubes all had different textures that could be felt. The haptic scene was rendered with the OpenHaptics renderer and all interface elements and objects were created by using nodes in the ReachIn API. The small and differently colored spheres shown in Figure 1 represent two users holding the same object. In this way the users can co-operate in compiling larger objects.

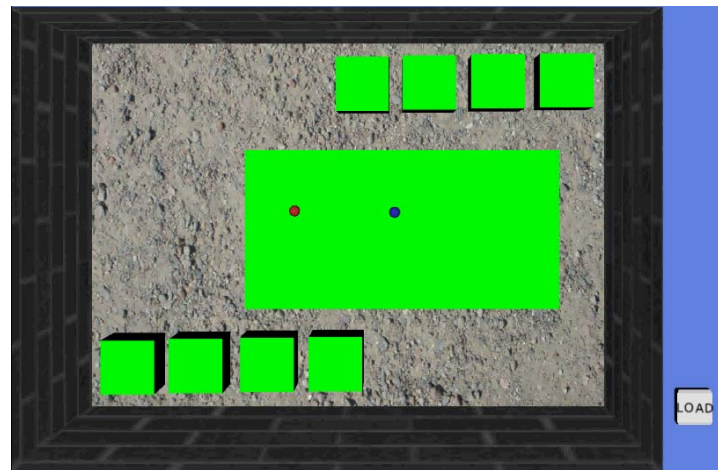


Fig. 1. The experimental collaborative interface including two users represented by one proxy each in the shape of one red and one blue sphere.

Since gravity and mass are applied to all objects, the users feel the weight and inertia of objects as they carry them around. Besides feeling and manipulating the cubes, users can feel and also grasp each other's graphical representations in order to provide navigational guidance e.g. to a blindfolded partner. The users can also "feel the other person's proxy" by means of a small vibration, applied whenever the users' graphical representations get close enough.

In the visual, haptic and audio interface, a number of auditory functions were added that gave different kinds of audio cues. A "grip sound" was heard every time an object was lifted. In this way the participant could detect that someone lifted an object. The second auditory function was a "put down sound" which was heard every time an object fell on the floor. In order to distinguish between the sound an object makes when it falls to the floor and the sound an object makes when it falls on another object, a "collision sound" was also designed which was heard every time an object landed on top of another one.



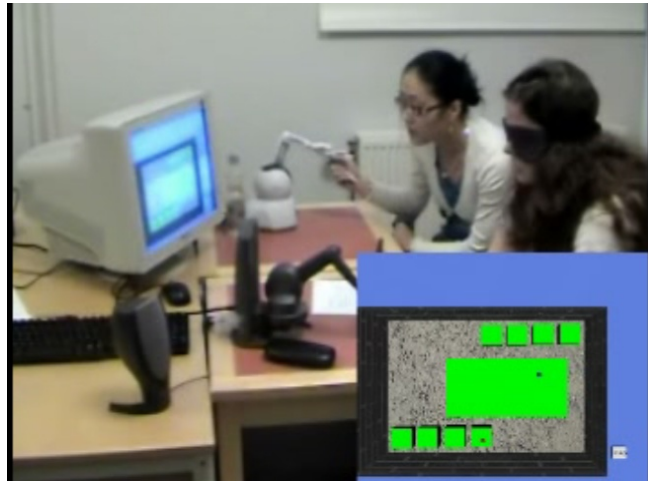


Fig. 2. The experimental setting with participants using one Phantom Desktop and one Phantom Omni respectively and a picture of the screen with spheres representing participants in the shared environment.

The fourth auditory function was a “contact sound” which was heard every time the button on the phantom was pushed down except, of course, when an object was grasped. This “contact sound”, in stereo, made it possible for a participant, and especially the blindfolded one, to know the location of the other user’s position relative to his/her own position. The loud speakers were placed on each side of the computer screen and as far out on the table as needed in order for both subjects to clearly hear that the sound came from the left or the right (Figure 2) relative to them. The sound appears to originate from the proxy but the sound environment is larger than the visual environment in order to make it possible for both subjects to distinguish if the sound is from the left or right side of them.

#### 4.4 Apparatus

The hardware used in this experiment were one personal computer with two dual core processors, a computer screen, a computer mouse, a keyboard, and a pair of loud speakers. Two different haptic devices were used; one Phantom Desktop (Figure 3) and one Phantom Omni (Figure 4). Both devices are operated with a pen-like stylus attached to a robotic arm that generates force feedback [Patomäki et al. 2004].



Fig. 3. A Phantom Desktop force feedback device (source: SensAble Technologies Inc).



Fig. 4. A Phantom Omni force feedback device (source: SensAble Technologies Inc).

Reachin API 4.1 and Microsoft Visual Studio 2003 .NET were used as software platform and CamStudio was used for screen capturing in order to record the interaction in the interface during the experiment. Ordinary loud speakers were used that were connected to the computer.

#### 4.5 Procedure

The researcher gave introductory information about the aim of the experiment, followed by an instruction on how to use the haptic devices. The researchers made sure that both participants were fully aware of how the haptic and audio feedback worked and could be utilized in the interface before the participants started solving the tasks.

The experiment was divided into four parts; one demo session, one training session, one group work session, and one interview session. In the demo session the soon to be blindfolded participant got the chance to use the Phantom Desktop while looking at the screen. In the demo environment he/she could feel several boxes with different textures and surfaces applied to them.

After one of the participants was blindfolded they then practiced on a training task in the experimental environment before the real task, in order to get used to this type of interface. They practiced how to feel the shape of a cube, how to navigate in the three-dimensional environment and how to grab a cube, lift it and hand it off to the other person in the group. It was made sure that the blindfolded participant felt comfortable with the blindfold and that she/he got used to working in this kind of haptic environment before the real experimental task was started. After the training session, each pair of participants solved a task in either the visual/haptic or the visual/haptic/audio condition. The blindfolded participant used a Phantom Desktop and the sighted used the Phantom Omni device in all sessions. Interviews were conducted after the task was performed in order to acquire information about system usability and subjective perception. In the interview open questions were asked to each participant of the pair.

#### 4.6 Task

The task performed in collaboration was to first move eight small cubes so that they formed four legs and then lift a larger flat rectangular object and place it on top of the legs so the result looked like a table. Collaboration was inevitable since they were only allowed to move four of the cubes each. However, they were allowed to collaborate in lifting the table board. The task was formulated as follows:

“In this assignment eight cubes of size 2 cm x 2 cm are placed on the floor, four at the bottom left and four at the upper right side of the room respectively. In the middle of the room there is also a large board with a size of 12 cm x 6 cm. Your assignment is to build a table. The task has been solved when the board has been positioned 4 cm above floor level. The table legs should be placed at the respective corners of the table. Before you start to solve this task you have to decide who is responsible for which group of cubes, you only have the right to move your own cubes.”

### 5 RESULTS

#### 5.1 Task Performance

One-way ANOVA (Analysis Of Variance) was used for analyzing the variable time. The hypothesis was that adding audio functions in the visual/haptic environment would improve task performance. The result showed that task performance, defined as the time to complete the task, differed significantly ( $p < 0.05$ ) across the two conditions. The mean value of the task completion time was shorter in the audio/haptic/visual condition ( $M = 581$  seconds,  $Sd = 284$ ) than in the haptic/visual condition ( $M = 916$ ,  $Sd = 277$ ), (Table 1). This means that subjects used about 10 minutes to perform the task in the audio/haptic/visual condition and about 15 minutes in the condition with no audio feedback.

**Table 1.** Experimental results regarding total time to complete task for 14 groups.

\*=significant at 95% level

The analysis of the audio cues especially to blindfolded	Task Performance (sec.)	Audio/Haptic Feedback Mean (Sd) n	Haptic Feedback Mean (Sd) n	conclusion quantitative that the provided, the person,
	F=5.0 p=0.045*	581 (284) 7	916 (277) 7	information about what actions the other person performed as well as the results of the actions the blindfolded person did in the environment. This is an important basic requirement that facilitates the grounding process, as the collaborators do not need to update each other about changes made in the environment. The qualitative analysis illustrates how these processes unfolded and validate results from the quantitative analysis.

## 5.2 Qualitative Analysis

The post-test interviews were transcribed word for word and a qualitative analysis was performed. A qualitative observation analysis was also performed of the video recorded material from the experimental sessions. The video recordings were transcribed word for word and notations of the behavior of the subjects were made.

The most important categories that were used in the analysis were awareness, common ground and usability. The definition of awareness in this study was “individuals’ perception of others’ activities and the status of the joint work-process”. In the interview and the observation analysis a number of aspects were considered like for example how the participants could maintain a shared focus of events in the joint task, how they perceived where the other person was located and how they could track movements of the other person continuously and know the status of ones’ own actions.

In this study common ground was defined as “a state of mutual understanding among conversational participants about the topic at hand” [Clark and Brennan, 1991]. In the interview and the observation analysis common ground was addressed using a number of categories that were, the participants’ shared understanding of the layout and their shared understanding of the objects’ qualities. Furthermore, the participants’ grounding strategies were analyzed.

Finally, how subjects could make use of different sensory modalities was analyzed and the extent to which sensory feedback helped them maintain a shared understanding of the work context. In the interviews and in the observation analysis a number of aspects were considered regarding usability. How well they understood the system and their perceived skill level in using specific features in the system were considered.

## 5.3 Result from Observations and Interviews

Here, a couple of dialog examples have been chosen in order to illustrate the findings regarding awareness, common ground and usability as related to the impacts of the different multimodal interfaces.

### 5.3.1 Awareness

Two aspects of awareness in focus in the qualitative analysis were how the blindfolded participant tracked changes in the environment and how well he/she knew the partner’s whereabouts and doings. The focus was also on how and with what modalities the sighted participant communicated information to the blindfolded participant in the two experimental conditions when performing joint tasks. The haptic feedback can be used to track changes while feeling around on the floor – one can feel if something has been moved or is showing up at a new place. The blindfolded participant then needs, however, to constantly explore the whole workspace, something that became evident when studying the groups who did not have access to audio feedback. In the audio groups, on the other hand, the sounds instantly inform the blindfolded participants if something has been moved. They do not need to explore actively to find out if something changed and they are always aware of the fact that work is in progress. Interestingly enough, one of the blindfolded participants in a sound group felt a little anxious

when she had not heard anything for a some time, and she asked her peer “So, what’s happening now?”. It had been silent for a while because the sighted was just looking at the situation without moving anything. In the groups without audio questions like “Have we started?” and “Are you doing anything?” were frequent, indicating that the blindfolded was unaware of that work was in progress.

Another advantage of audio cues is that they give the blindfolded participants a confirmation of what they are doing themselves. It was clear from our observations that the force feedback generated when lifting an object did not give enough information to the blindfolded participant. The same could be said about the dropping of a cube. In our haptic only groups, questions like “Did I drop it?” and “Did I pick it up?” were frequent. Obviously, the touch feedback was not enough to confirm the actions. These kinds of questions were never asked in the audio groups. The sound was enough confirmation of that the blindfolded participants did what they intended to do. It would definitely be possible to do things without audio feedback, but the sense of control would probably be affected in a negative way. The example below shows yet another advantage of the collision sound.

[Blindfolded accidentally drops the cube she is holding]

[The collision sound is heard]

Blindfolded: Oh, I dropped it now

Sighted: Yes, you can pick it up again if you feel it

[Blindfolded picks up the cube again]

In one case the blindfolded dropped a cube several times in a row without noticing it. Each and every time he had to wait for the sighted to tell him that the cube was dropped.

The difference between the two put down sounds did also convey valuable information, despite the fact that they were pretty similar. This is illustrated in the following example, where the blindfolded is placing a cube on another cube to build a leg.

[Blindfolded puts the cube down, floor collision heard]

[Blindfolded puts it down again, floor collision again]

[Blindfolded lifts and places the cube on another, object collision heard]

Blindfolded: Does it look nice?

Sighted: No...

Blindfolded: But I know it’s on top...

In this case the sighted never had to direct the blindfolded for her to be able to place the cube on another one – the sound was enough. In the haptic only groups the sighted had to give verbal guidance while the blindfolded was moving up in the air. It is easy to see that sound makes a big difference for the better in this case. The blindfolded participant could actually do some things completely by him/herself – he/she is not completely dependent on the sighted peer.

That sound increases the understanding of the workflow in our application is fairly obvious. It also increases the work efficiency, providing information that otherwise had to be conveyed through verbal guidance and communication. The sound model gives the blindfolded participant

a sense of what the other person is doing and that work is in progress as well as confirmation of his/her own actions.

Furthermore, the grip and contact sounds made the blindfolded participants aware of what the sighted did and approximately where he/she was. All participants in our audio groups preferred to work in a virtual environment with audio feedback, reflecting their actions, rather than in silence. This was highlighted in the interviews with the sighted and blindfolded participants in these groups.

The blindfolded participants in the haptic-only groups, however, complained about not knowing what was going on. They felt that they were working in silence and that they could not contribute to the work process. If sound is not available the only way for the blindfolded to be aware of what is going on is by talking to the sighted peer, who cannot possibly tell every detail about what is happening.

Furthermore, the blindfolded participants in the haptic-only groups did not get any kind of audio feedback on their own actions – they had to ask their sighted peers if they had dropped an object, since the touch feedback apparently was not always enough.

Based on this discussion and the above example we argue that the audio cues make a considerable positive difference regarding the amount of awareness information accessible for the blindfolded and sighted participants in the collaborative work.

### 5.3.2 Common Ground

Another result from the observations was that our audio groups could use the audio feedback as an aid when establishing common ground.

Besides using the audio cues to pick up and put down objects, the sighted participant often used the contact sound to show where a certain cube could be found and placed respectively.

Below is a typical conversation that took place when a group tried to solve a subtask on placing one cube on another. In the example, the potential of the added audio feedback for deictical referencing during navigation in a haptic interface is made clear. Deictic reference to objects is a fundamental grounding strategy.

Sighted: Pick up a new cube!

[The sighted used contact sound to express where a cube was and the Blindfolded located it thanks to the sound]

Blindfolded: That one?

[The Blindfolded picked up the cube, and a grip sound occurred]

Sighted: Yeah...And then you can move here...

[Sighted uses the contact sound to show the way again]

[Blindfolded navigates to a place slightly above the intended one]

Sighted: Ok, down a bit..., down..., and stop.

[The sighted tries to help the Blindfolded to navigate, verbally]

Sighted: Ok, drop it.

[The Blindfolded releases the cube on another one, and hears a collision sound]

Sighted: You can try to pick up the cube that's here...

[Sighted uses the contact sound again]

[Blindfolded navigates to the exact place in a few seconds]

This is an example of how the sighted user manages to get the blindfolded person's attention and guide the blindfolded to a certain place. We can also see from the example that the blindfolded ended up at the right place quickly. In the groups lacking the audio feedback, the sighted participant often had to spend a lot of time explaining in words (like “up”, “down”, “too much, go back”) in what direction to move and the blindfolded had to go back and forth numerous times before ending up at the intended place. When the sighted do not need to verbally direct the blindfolded, they can instead focus on describing the different objects' shapes – thus, the dialog becomes more goal-oriented.

In most of our audio groups the sighted participant used speech and the contact sound in combination to describe the workspace's layout and the objects in it. This provided them both with a common ground of the shared workspace.

The example above illustrates another advantage given by the audio feedback, that sounds give the sighted participant important information as well. He tells the blindfolded to grasp a certain cube, he waits until he hears the grasp sound after which he starts to guide. In the haptic only groups the sighted participant had to ask questions like “Did you drop it?” or “Did you grasp it?” all the time. Since those kinds of questions were never asked in the audio groups we can conclude that the sound gives enough confirmation and that it gives valuable information to both the blindfolded and the sighted participant. Thus, the sound simplifies the collaboration – the blindfolded participants can concentrate more on the task than on the constant need to ask and in other ways find out what happens. In most of our haptic/visual groups the sighted had to devote a lot of time to guide the blindfolded peer verbally, something that was often cumbersome.

In the interviews with the participants in the audio groups most of the sighted ones highlighted the importance of being able to achieve a common ground by means of audio and haptic feedback in the task solving process. They also appreciated the sound feedback when it came to making the blindfolded aware of what was happening. The blindfolded participants could use the audio feedback as an aid when navigating and orientating and they felt that it was possible to feel things with the haptic feedback equipments.

### 5.3.3 Usability

Interestingly, the fact that the blindfolded participants were unable to see the cubes on the screen made it possible for them to feel the size of cubes quite easily, while the sighted ones, that sometimes disregarded the sense of touch before making a judgment, had a harder time estimating the size. As long as the participants established a common ground by talking, navigating and exploring the workspace, they could reach the same level of understanding of the layout, the context and task at hand.

Many of the participants in our haptic-only groups said that the addition of sound feedback would be a great aid in their collaboration.

The findings in this study showed the importance of the audio feedback when it came to supporting collaboration, although the haptic feedback was an important prerequisite. That is consistent with results from several studies that have shown the importance of the haptic modality in collaborative environments. Our study is no exception, the haptic feedback made it

possible for the blindfolded to feel objects enabling for both participants to focus their attention on and talk about the same object as if they really perceived that they were working in the same workspace with the same work material. Utterances about specific places in the environment and different sizes of object show that the participants share a common view about the properties of the objects in the virtual environment. The haptic feedback is making this possible.

When comparing the two experimental conditions, we identified a positive qualitative difference caused by the addition of sound cues. In the audio and haptic condition, the blindfolded solved the tasks and navigated not only using their sense of touch combined with the verbal direction from the sighted participant, but also their sense of hearing. The audio modality made it possible for them to orientate (with a contact sound), to hear approximately what their partner was doing and where he/she was and to hear what they did themselves. The fact that the sighted person did use the contact sound and the blindfolded did interpret this sound cue as being the sighted person's position in the virtual environment and could move in the right direction in response to it, implies that this sound might be an example of a feature that shortened the time a blindfolded person spent navigating which in turn made task performance faster.

In the haptic/visual groups, some blindfolded participants said that they had no idea whether they did something (such as dropping an object on another one or dropping an object on the floor) or not. Moreover, they did not know what their partner did. Almost all participants in the haptic/visual groups believed that it would make a positive difference if audio was added in the environment. The blindfolded participants in these groups wanted to know that something was going on, and to get confirmations on actions taken.

## 6 DISCUSSION

In this paper we have shown that audio feedback can be utilized by groups of one sighted and one blindfolded person in order to collaborate more efficiently in a virtual haptic environment.

The data from this experiment showed that the interface, providing audio as well as haptic feedback, made the groups perform the task significantly faster compared to when audio cues were not provided. The result supports the hypothesis that collaboration would take less time when audio cues provide information about the actions that the two participants made in the haptic virtual environment. According to McLeod [1996], the usability of a system can be measured by how long it takes to perform a task and how well the task is performed. It has been shown that the larger the number of modalities, the higher performance would be expected from the participants [Short et al. 1976].

The qualitative results explain in what ways the auditory cues play an important role for collaboration in multimodal environments. The added audio information improved the possibility to get awareness of actions and to locate the other person in the virtual environment. This in turn supported the grounding process when participants solved a task collaboratively. This is in accordance with the common ground theory that states that actions can be communicative and that actions are consciously used to for example direct others' attention to objects or places [Clark, 2003]. In order to be communicative, actions have to be possible to perceive. In ordinary life actions are seldom soundless. The fact that things make sounds when moved or when people interact with them is, apart from visual feedback, useful to take into account in a virtual environment, as action cues. In a collaborative situation these sounds provide



an added value when coordinating actions of several people as well as their communication. Several researchers have emphasized the importance of providing continuous awareness of others' actions and the status of others' work processes in a virtual environment [Dourish and Bellotti 1992; Benford et al. 1994]. It is a challenge however, to decide what sounds to include and what purposes they should fulfill, since the sound environment could easily become too complex. This was the reason why only a few fairly simple interaction sounds were used in this experimental study. In this study, the sounds such as the "grip sound", that was heard every time an object was lifted, the "put down sound", that was heard every time an object touched the floor and the "collision sound", that was heard every time an object landed on top of another object, were designed in a very simplistic fashion. These kinds of sounds were designed only as event oriented cues of actions taken. They did not give any advanced information about the work process. An example of an alternative approach that might be pursued in future studies is using continuous friction sounds, heard whenever a user touches and manipulates an object or surface.

It might also be the case that some other type of haptic feedback could provide better awareness of the manipulations of objects by the other person as well as actions made by oneself. However, when a person manipulates an object she gets haptic feedback of the object's surface and its weight and in that situation audio adds more information. The manipulation of objects by the other person is hard to represent better by haptic feedback and here audio cues are clearly providing additional and unique information. The action awareness gained through all the audio cues together makes it easier to maintain a common ground of the state of the task performance process. It should be noted that the audio cues are most probably useful not only for the blindfolded but for the sighted person too even if it is of more obvious importance to the blindfolded.

The findings in this study are important to consider when designing multimodal interfaces for collaboration between sighted and visually impaired people in order to make collaborative situations more accessible to visually impaired people. In view of the principles of design for communication and collaboration between sighted and blindfolded people [Winberg and Bowers 2004], the results suggest that the auditory/visual/haptic interface of our prototype is functional. All the basic actions that could be done by sighted users of the GUI could be done by the blindfolded as well. The auditory and haptic feedback enabled manipulation and exploration in the shared virtual space and that feedback was coherent with the visual interface, which enabled collaboration between sighted and blindfolded participants.

In their study of an auditive "Towers of Hanoi" game Winberg and Bowers [2004] showed that visually impaired persons could perform the task of moving discs from one peg to another together with a sighted person and that they could have a dialogue about the shared workspace. In that environment no haptic feedback was provided. In our study we have verified that audio cues significantly improve efficiency compared to a haptic environment without sound.

We argue that the results from the study presented here can be generalized to groups with visually impaired people included despite the fact that we used blindfolded sighted people. The hypothesis tested in this experiment was based on results from an earlier field study of group work between sighted and visually impaired pupils in primary school using a haptic and visual interface [Sallnäs et al. 2007]. The haptic and visual interface used in the previous field study was the same, regarding the functionality, as the ones in the experiment presented here, except that audio cues were added. Therefore, the application has been validated with visually impaired

persons on all accounts except the audio cues. The next step however, would certainly be to test the application with visually impaired persons now when the usefulness of the audio cues has been established for collaboration between sighted and blindfolded persons.

#### ACKNOWLEDGMENTS

This study is part of the project Multimodal Collaborative Environments (MCE) funded by the Swedish Research Council (VR) and the EUfunded project MICOLE. Many thanks go to Professor Gunnar Jansson for his warmth in discussions and suggestions and to the participants in the experiments.

#### REFERENCES

- BARFIELD, W. AND FURNESS, T. A., EDS. 1995. *Virtual Environments and Advanced Interface Design*. Oxford University Press, Inc., New York, NY.
- BASDOGAN, C., HO C., SRINIVASAN, M.A., AND SLATER, M. 2000. "An Experimental Study on the Role of Touch in Shared Virtual Environments", *ACM Transaction of Computer Human Interaction*, 7(4), 443–460.
- BASDOGAN, C., HO, C., SLATER, M., AND SRINIVASAN, M. A. 1998. The role of haptic communication in shared virtual environments, in *Proceedings of the Third Workshop on Phantom User Group* (Dedham, MA, Oct. 3-6). <http://www.ai.mit.edu/conferences/pug98>.
- BENFORD, S., BOWERS, J., FAHLÉN L. E., AND GREENHALGH C. 1994. Managing mutual awareness in collaborative virtual environments. In *proceeding of Virtual Reality Software and Technology*, Singapore, 223–236.
- BERGAMASCO, M. AVIZZANO, C. DI PETRI, G., BARBAGLI, F. AND FRISOLI, A. 2001. The museum of pure form: system architecture, In *Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication*, Piscataway, NY: IEEE Press, 112-117.
- BERGAMASCO, M. AND PRISCO, G. 1998. Design of an anthropomorphic haptic interface for the human arm, In Y. Shirai and S. Hirse (Eds.), *Robotic Research, the Eight International Symposiums*, 278-289. London: Springer.
- BURKE, J., AND MURPHY, R. 2007. RSPV: An Investigation of Remote Shared Visual Presence as Common Ground for Human-Robot Teams. In *Proceedings of HRI'07*, March 8-11 2007. Arlington, Virginia, USA. 161-168.
- CARROLL, J.M., NEALE, D.C., ISENHOUR, P.L., ROSSEN,M.B., AND MCCRICKARD, D.S. 2003. Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human Computer Studies*, 58, 605-632.

- CARROLL, J.M., ROSSON, M.B., CONVERTINO, G., AND GANOE, C.H. 2006. Awareness and teamwork in computersupported collaborations, *Journal of Interacting with Computers*, 18(1), 21-46.
- CHERUBINI, M., NUSSLI, M-A., AND DILLENBOURG, P. 2008. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of ETRA 2008*, ACM, New York, 173-180.
- CLARK, H. H. 2003. Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243-268). Hillsdale NJ: Erlbaum.
- CLARK, H. H. 1996. *Using Language*, New York, Cambridge University Press.
- CLARK, H. H. AND BRENNAN, S.E. 1991. Grounding in Communication. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition*, Hyattsville, MD: American Psychological Association, 127-149.
- CROMMENTUIJN, K. 2006. Designing Auditory Displays to Facilitate Object Localization in Virtual Haptic 3D Environments, In *Proceedings of 8th ACM Conference on Computers and Assessibility*, October 22-25, 2006, Portland, Oregon, USA, 255 – 256.
- DOURISH, P., AND BELLOTTI, V. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 541-547.
- DURLASH, N., AND SLATER, M. 2000. Presence in shared virtual environments and virtual togetherness. *Journal of Presence: Teleoperators and Virtual Environments*, 9(2), 214-217.
- ELLIS, S. R. 1992. Nature and origin of virtual environments: A bibliographic essay. *Comput, Syst. Engin.* 2(4), 321–347.
- FRISOLI, A., SIMONCINI, F. AND BERGAMASCO, M. 2002, Mechanical design of a haptic interface for the hand, In *Proceedings of 2002 ASME DETC 27th Biennial Mechanisms and Robotics Conference*, Montreal, Canada: ASME Press, 25-32.
- FUSSELL, S.R., SETLOCK, L.D., YANG, J., OU, J., MAUER, E.M., AND KRAMER, A. 2004. Gestures over video streams to support remote collaboration on physical tasks. In *Human-Computer Interaction*, 19, 273-309.
- GUPTA, R., SHERIDAN, T. AND WHITNEY, D. 1997. Experiments Using Multimodal Virtual Environments in Design for Assembly Analysis. *Presence: Teleoperators and Virtual Environments*. 6(3), 318-338.

- HASSER, C. J. GOLDENBERG, A. S. , MARTIN, K. M. AND ROSENBERG, L. B. 1998. User Performing a GUI Pointing Task with a Low-Cost Force-Feedback Computer Mouse, DSCVol. 64, In Proceedings of the ASME Dynamics and Control Division, 151-156.
- HELD, R. M. AND DURLACH, N. I., 1992, Telepresence, Presence: Teleoper. Virtual Environment. 1(1), 109-112.
- HO, C., BASDOGAN, C., SLATER, M., DURLACH, N., AND SRINIVASAN, M. A. 1998. An experiment on the influence of haptic communication on the sense of being together, In Proceedings of the British Telecom Workshop on Presence in Shared Virtual Environments (Ipswich, Suffolk, June 10–11).
- HURMUZLU, Y., EPHANOV, A. AND STOIANOVICI, D. 1998. Effect of a Pneumatically Driven Haptic Interface on the Percept ional Capabilities of Human Operators. Presence: Teleoperators and Virtual Environments. 7(3), 290-307.
- JANSSON, G. 2007. Haptics as a substitute for vision, in M. A. Hersh and M. A. Johnson (Eds.), Assistive technology for visionimpaired and blind, London: Springer.
- KENNEL, A., AUDIOGRAF. 1996. A diagram reader for blind people, In Proceedings of ASSETS'96, Second Annual ACM Conference on Assistive Technologies, Vancouver: ACM Press, 51-56.
- KIMMERLE, J., CRESS, U., AND HESSE. F.W. 2007. An interact ional perspective on group awareness: Alleviating the information-exchange dilemma (for everybody?). International Journal of Human-Computer study, 65(11), 899-910.
- KIRK, D., RODDEN, T., AND FRASER, D. S. 2007. Turn it This Way: Grounding Collaborative Action with Remote Gestures. In Proceedings of CHI 2007, April 28-May 3. San Jose, California, USA, 1039-1048.
- KRAUT, R.E., FISH, R.S., ROOT, R.W., AND 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, 287-314. San Mateo, CA:Kaufman.
- KUBER, R., YU, W., AND MCALLISTER, G. 2007. Towards Developing Assistive Haptic Feedback for Visually Impaired Internet Users, In Proceedings of the 2007 Conference on Human Factors in Computing Systems, San Jose, California, USA, 1525 – 1534.
- LOOMIS, J. M., AND LEDERMAN, S. J. 1986. Tactual perception, In K. R BOFF, L. KAUFMAN & J.P THOMAS (Eds.). Handbook of perception and human performance, Wiley/Interscience, New York, 31.31-31.41.

- MCCARTHY, J., MILES, V., AND MONK, A. 1991. An experimental study of common ground in text-based communication. In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (209-215). New Orleans, Louisiana, US, New York, NY: ACM Press.
- MCGOOKIN, D. AND BREWSTER, S. 2007 An Initial Investigation into Non-Visual Computer Supported Collaboration, In Proceeding of CHI 2007, April 28-May 3, 2007 • San Jose, CA, USA
- MCLEOD, P. L. 1996. New communication technologies for group decision making: Toward an integrated framework. In Communication and Group Decision Making, R. Y. Hirokawa and M. S. Poole, Eds. Sage Publications, Inc., Thousand Oaks, CA, 426–461.
- NEALE, D.C., CARROLL, J.M., AND ROSSON, M.B. 2004. Evaluating Computer-Supported Cooperative Work: Models and Frameworks, In Proceedings of the International Conference on Computer-Supported Cooperative work , November 6-10, 2004, Chicago, Illinois, USA, 112-121.
- OAKLEY, I., BREWSTER, S. AND P.D. GRAY. 2001. Can You Feel the Force? An Investigation of Haptic Collaboration in Shared Editors. In Proceedings of Eurohaptics 2001, 54-59.
- OU, J., FUSSELL, S. R., CHEN, X., SETLOCK, L. D., AND YANG, J. 2003. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In Proceedings of ICMI 2003. ACM, New York, NY, 242-249.
- PATOMÄKI, S., RAISAMO, R., SALO, J., PASTO, V. AND HIPPUA, A. 2004. Experiences on Haptic Interfaces for Visually Impaired Young Children. In Proceeding of the Sixth International Conference on Multimodal Interfaces (ICMI'04), October 13–15. State College, Pennsylvania, USA, 281-288.
- PLIMMER, B., CROSSAN, A., BREWSTER, S. AND BLAGOJEVIC, R. 2008. Multimodal Collaborative Handwriting Training for Visually-Impaired People. In Proceedings of CHI '08. ACM Press, New York, NY, 393-402.
- POLL, L.H. AND EGGEN, B.H., 1996. Non-visual Interaction with GUI Objects, In Proceedings of HCI on People and Computers XI, 159-168.
- SALLNÄS, E-L., RASSMUS-GRÖHN, K. AND 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., MOLL, J. AND SEVERINSON-EKLUNDH, K. 2007. Group Work about Geometrical Concepts Including Blind and Sighted Pupils, In proceedings of World Haptics 2007, Tsukuba, Japan, 330-335.
- SCHMIDT, K. 2002. The Problem with 'Awareness', Introductory Remarks on 'Awareness in CSCW' Computer Supported Cooperative Work 11: 285–298, 2002.
- SHORT, J. WILLIAMS, E., AND CHRISTIE, B. 1976. The social psychology of telecommunications. London: Wiley. SJOSTROM, C. 2001. Designing Haptic Computer Interfaces for Blind People. In Proceedings of ISSPA, 1, 68-71.
- SLAVIN, R.E. AND COOPER, R. 1999. Improving Intergroup Relations: Lessons Learned from Cooperative Learning Programs. *Journal of Social Issues*. 55, 4, 647-663.
- WINBERG, F., AND HELLSTRÖM, S-O. 2001. Qualitative aspects of auditory direct manipulation, In Proceedings of CUU '00, ACM Press, New York, NY, 16-20.
- WINBERG, F., AND BOWERS, J. 2004. Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users, In Proceedings of CSCW'04, Chicago, Illinois, USA, 332-341.