The Influence of Modality Combinations on Communication in Collaborative Virtual Environments

JONAS MOLL

Avhandling som med tillstånd av Kungliga Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen torsdagen den 28 november 2013 kl. 13:00 i sal V2, Teknikringen 76, Kungliga Tekniska Högskolan, Stockholm.
Abstract

Although many studies have been performed on collaboration in multimodal interfaces not many of these have looked specifically on how the supported modalities influence the task solving strategies chosen and the communication between users solving a joint task in collaborative virtual environments. Therefore, the thesis studies performed aimed at shedding light on these aspects of multimodality. The specific research question studied is: How do changes in modality combinations influence employed work strategies, communication during task solving and the task efficiency in collaborative multimodal virtual environments? The studies performed build on theories from HCI, CSCW, human perception and mediated communication and are thus inter-disciplinary in nature. A variety of cases have been studied; collaboration between sighted and visually impaired, task solving in visually demanding environments and to some extent support for achieving medical diagnoses.

The research presented in this thesis began with a field study in elementary schools, focusing on collaboration between visually impaired and sighted pupils. The shared environment was in this case a virtual room in which objects could be moved around by means of haptic devices. The results showed a great potential for haptic feedback when it came to supporting collaboration and most of all communication between the participants. A lack of awareness information about mostly the sighted pupils’ actions laid the ground for a follow-up study in which sighted and blindfolded students solved tasks in the same interface. A formal experiment was carried out in this case, comparing a visual/haptic environment with a visual/haptic/audio environment. Results showed that the addition of audio feedback to the visual/haptic environment was beneficial in many respects. Up until now, the focus had been entirely on collaboration between sighted persons and those who cannot see. This is why the next experimental study, based on an abstract gaming environment, aimed at collaboration between sighted persons. Since the earlier studies showed that the combination of modalities clearly matter, this new experiment compared three modality combinations – visual/haptic, visual/audio and visual/haptic/audio. Once again, the results clearly showed that the combination of modalities has an effect on task performance and that it influences collaboration and communication in particular.

All studies performed have been subject to both quantitative analysis of performance measures and qualitative analysis of dialogues between collaborators. Even though quantitative data on task performance has played an important role, the main focus has been on qualitative data in all studies performed. The results show that different combinations of modalities influence the collaboration and in particular the communication between two participants solving tasks in different ways in a number of multimodal interfaces. In all cases in which a visual/haptic/audio condition has been compared to a visual/haptic or a visual/audio condition the performance was significantly better in the visual/haptic/audio condition. One of the most important conclusions drawn from the qualitative analysis of dialogues is that both haptic and audio feedback can have communicative properties which influence the dialogue and as a consequence the collaboration.
Foreword

When I, as a computer science student at KTH, first came in contact with haptic virtual environments during a guest lecture I felt that I wanted to explore that very special research area in more depth. I knew nothing about the area from before and at that point I had no idea that the lecturer, Eva-Lotta Sallnäs Pysander, would become one of my supervisors in the most challenging project I have ever been a part of – the one resulting in this thesis – about four years later.

I have always been interested in computer support for communication and thus it felt natural to focus on collaboration in multimodal virtual environments and especially on the communicative aspects of haptic feedback. Even though haptic feedback has always been my main interest, audio feedback has received more and more attention as the work has evolved. Overall, I’m very pleased with the outcome of the work I have performed during these five years – not only did I get a chance to dig deep into how haptic and audio feedback can affect communication and collaboration in virtual environment (an under-investigated research area) but I also developed my skills as a researcher by working more and more independently with studies and publications as work evolved. The only regret I have is that I did not have the opportunity to investigate the role of haptic feedback in the area of medicine in more depth – the area is important in many respects and all attempts to support communication can be literally life altering. Anyway, it is always good to have something to build on for future work.

It goes without saying that I could not have done everything reported in this thesis entirely on my own. First of all, I have received great help from my supervisors Eva-Lotta Sallnäs Pysander and Kerstin Severinson Eklundh. Without their support throughout the process I could never have pulled this off. We have always had very fruitful and constructive discussions which have strengthened my abilities to do research and communicate it through scientific publications. Second, the work would have been much harder to perform and I would have been a lot more self-absorbed if it had not been for the collaboration with other research students working in parallel with me in joint projects. Most and foremost I am thinking about Yingying Huang, with whom I planned and conducted an entire study, Sten-Olof Hellström, who had the skills for developing sound models which I did not have and Oscar Frykholm and Jonas Forsslund, who I collaborated with in the medical project Funk-IS. Last, but not least, the work would not have been possible to conduct without all participants taking part in the studies. My thanks goes out to all pupils, students and surgeons who made the choice to devote an hour of their lives to my strange projects!
## Table of Contents

### 1 Introduction 1

1.1 Research focus and question 2

1.2 The thesis studies – short overview 2

1.3 Outline 4

### 2 Theoretical background and related research 6

2.1 Working with multimodal interfaces 6

2.1.1 Vision, hearing and touch in real and virtual environments 7

2.1.2 Cross-modality relations 17

2.1.3 Benefits of multimodal input 19

2.2 Collaborating in virtual environments 22

2.2.1 Important terms and definitions 23

2.2.2 Awareness and common ground in this thesis 27

2.2.3 Collaborative multimodal virtual environments 28

2.3 Communicating in multimodal virtual environments 34

2.3.1 Touch in human communication 34

2.3.2 Communicating by mediated touch 35

2.3.3 Haptic feedback can influence the dialogue 40

### 3 Methodological concerns 41

3.1 Analyzing dialogues 41

3.1.1 Dialogue analysis approaches 41

3.1.2 Dialogue analyses in this thesis 42

3.2 Quantitative and/or qualitative data? 44

### 4 Haptic feedback for collaboration between visually impaired and sighted pupils 46

4.1 Group work in Swedish elementary schools 47

4.1.1 Some theoretical points about learning 47

4.1.2 Differences in work material 48

4.2 Software support for visually impaired 49
4.2.1 Designing haptic interfaces for visually impaired users ............................................. 49
4.2.2 Multimodal environments for visually impaired users .............................................. 50
4.3 Developing new interfaces ............................................................................................ 54
  4.3.1 The static application ................................................................................................. 56
  4.3.2 The dynamic application ............................................................................................ 57
4.4 Evaluation in the field .................................................................................................... 58
  4.4.1 Subjects .................................................................................................................. 59
  4.4.2 Apparatus ................................................................................................................. 59
  4.4.3 Setting ...................................................................................................................... 59
  4.4.4 Procedure and tasks ................................................................................................. 60
4.5 Analyses ........................................................................................................................ 62
  4.5.1 Quantitative measures ............................................................................................... 63
  4.5.2 Qualitative measures ................................................................................................. 63
4.6 Main results ................................................................................................................... 64
  4.6.1 Haptic feedback’s influence on verbal communication ............................................ 65
  4.6.2 Providing guidance ................................................................................................... 66
  4.6.3 The static vs. dynamic environment ......................................................................... 69
4.7 Lessons learned ............................................................................................................. 71
  4.7.1 Communicating with haptic feedback ....................................................................... 71
  4.7.2 The importance of awareness feedback .................................................................... 74

5 Haptic and audio feedback for collaboration between sighted and blindfolded adults ......................................................... 76
  5.1 Incorporating sound feedback in a haptic interface ...................................................... 77
  5.2 The experiment ............................................................................................................ 78
    5.2.1 Subjects ................................................................................................................. 79
    5.2.2 Apparatus .............................................................................................................. 80
    5.2.3 Setting .................................................................................................................... 81
    5.2.4 Procedure and tasks ............................................................................................ 81
  5.3 Analyses ........................................................................................................................ 83
    5.3.1 Quantitative measures ............................................................................................ 84
    5.3.2 Qualitative measures ............................................................................................. 85
  5.4 Main results .................................................................................................................. 85
5.4.1 Quantitative results ...................................................................................... 86
5.4.2 Qualitative results ...................................................................................... 87
5.5 Lessons learned ........................................................................................... 93
  5.5.1 Deepened insight into haptic communication ........................................... 93
  5.5.2 The difficulty and necessity of providing accurate feedback ..................... 95
  5.5.3 The audio could make task performance faster .......................................... 97

6 Scanning and placing in an abstract multimodal interface .......................... 98
  6.1 Scanning for and placing objects in multimodal interfaces ....................... 99
  6.2 Developing the interface ............................................................................. 100
  6.3 The experiment ........................................................................................ 102
    6.3.1 Subjects .................................................................................................. 103
    6.3.2 Apparatus .............................................................................................. 103
    6.3.3 Setting .................................................................................................... 103
    6.3.4 Procedure and tasks ............................................................................... 104
  6.4 Analyses .................................................................................................... 105
    6.4.1 Quantitative measures ........................................................................... 105
    6.4.2 Qualitative measures ............................................................................ 106
  6.5 Main results ............................................................................................. 107
    6.5.1 Quantitative results ................................................................................ 107
    6.5.2 Qualitative results .................................................................................. 109
  6.6 Lessons learned ......................................................................................... 114
    6.6.1 About scanning and placing strategies ................................................... 114
    6.6.2 Influences on group dynamics ............................................................... 115
    6.6.3 Sound and haptic feedback for emptiness .............................................. 117

7 Discussion of contributions ......................................................................... 120
  7.1 The combination of modalities matters ...................................................... 120
  7.2 Multimodal communication revisited ......................................................... 122
    7.2.1 Specific functions for communication .................................................... 122
    7.2.2 Influences on the dialogue ..................................................................... 124
  7.3 Technical considerations ............................................................................ 126
  7.4 Ongoing research on haptics within the medical domain .......................... 128
7.5 Some reflections on generalizability ................................................................. 132

References ............................................................................................................. 134
1 Introduction

Researchers in many different fields have studied how the human senses work and how we can use our knowledge about the senses to develop usable and effective computer interfaces. Our sense of vision and its implication for graphical design is by far the most researched area in this respect. Studies on visual dominance, gestalt principles and pre-attentive processing, just to name a few areas, have laid the ground work for many successful applications.

Our sense of hearing has also undergone many studies and several examples can be found in which audio is successfully used to e.g. convey warning and status messages and there are even examples of interfaces based solely on audio (without an accompanying visual display).

During the last few decades researchers have also taken an increasing interest into how our sense of touch can be utilized for interaction with haptic computer interfaces. Many different kinds of haptic input equipment have been developed and the most widely known are the PHANTOM devices. These devices use pen-like styluses attached to robotic arms, which generate contact forces when the proxy\(^1\) coincides with a graphical object. Interaction in haptic interfaces is still a relatively new research area and it is constantly growing due to the potential for haptic feedback in e.g. medical training, data analysis and support for visually impaired computer users.

When it comes to the research area of computer supported collaborative work (CSCW) vision and audio have, yet again, received most attention. However, during the last decade haptic collaborative environments have been investigated as well. The focus of these studies has often been on how haptic and/or audio feedback can be used to allow joint task solving and on how haptic and/or audio feedback can be used to achieve awareness and common ground in the collaborative process. Means of communicating through mediated technology have been studied in depth when it comes to visual interfaces (e.g. solutions for video conferencing) but this is not the case for interfaces supporting other modalities. To be able to communicate through mediated technologies, no matter what the modality, common ground about messages sent is necessary for efficient collaboration. Even though there are established symbolic meanings of touch (Haans and IJsselsteijn, 2006), these are not generally agreed upon and the role of touch for communication in haptic interfaces is under-investigated.

Studies have been conducted on interaction in multimodal virtual environments for decades and several benefits of combining different modalities have been identified. Among the more pronounced benefits are that cognitive load decreases (Garcia et al., 2009), task performance is improved (Otaduy and Lin, 2005; Garcia et al., 2009; Sallnäs 2004) and the amount of awareness information about the overall work state is increased (Garcia et al., 2009; Sallnäs, 2004) when more senses are added to a visual interface. These results concern both single-user and multi-user environments.

Although many studies have been performed on collaboration in multimodal interfaces not many of these have looked specifically on how the supported modalities influence the work strategies chosen and the communication between users solving a

\(^1\) In this thesis proxy refers to the haptic device’s graphical representation in the virtual environment.
joint task in a collaborative virtual environment (CVE). Therefore, the thesis studies performed aim at shedding light on these aspects of multimodality and hence my main contributions to the research on collaboration in multimodal interfaces can be found there. The focus is on how haptic and audio functions can be used for communicative purposes during joint task solving in multimodal interfaces, and on how such functions influence the dialogue between the collaborators.

1.1 Research focus and question

The focus of all studies carried out within the scope of this thesis is on collaborative work in multimodal interfaces and especially on how different combinations of modalities influence communication and task solving strategies. Even though the thesis studies, introduced in section 1.2 below, concern different application areas the focus in all of these studies is on the collaborative process and how different modality combinations influence the communication.

Both quantitative and qualitative data have been gathered and analyzed, although the emphasis has been on the qualitative data. The focus in the quantitative analyses has been on how changes in modality combinations affect task performance when it comes to time to perform tasks and errors made. The qualitative focus has been on content analyses of dialogues between participants performing tasks together.

The work in all thesis studies has been guided by the following general research question:

*How do changes in modality combinations influence employed work strategies, communication during task solving and the task efficiency in collaborative multimodal virtual environments?*

The studies performed build on theories from HCI, CSCW, human perception and mediated communication and are thus interdisciplinary in nature. Next, these studies will be briefly introduced.

1.2 The thesis studies – short overview

The studies performed within the scope of the thesis work have been carried out in different research areas. Note that the focus will not be on the specific areas presented here, but on the aspects mentioned in section 1.1 above. In this section, the studies performed will be briefly introduced.

*Multimodal interfaces for collaboration between sighted and non-sighted*

Two of the thesis studies performed are based on interfaces, developed within the thesis work, for collaboration between sighted and non-sighted persons. The first of these studies focused on collaboration between visually impaired and sighted pupils in elementary school (see figure 1). An evaluation was carried out on two applications (one static for feeling drawn shapes and one dynamic in which objects could be grasped and
moved around by means of haptic feedback).
Groups of three pupils, of which one was visually impaired, solved a few simple geometrical assignments together by means of haptic feedback equipment. The rationale behind this study was, first and foremost, to see if access to a shared interface would ease the inclusion of the visually impaired pupil in group work. How the haptic feedback was used to create a common frame of reference (common ground) was also of great importance here. This is also the first of the thesis studies in which haptic guiding functions were put to use. A haptic guiding function, which will be defined more formally later on in the thesis, is a haptic function that enables one user to drag the other one to a certain place through e.g. a magnetic force between the two proxies. This function enables close to physical interaction in real time.

The focus in chapter 4, covering this study, is quite different from the focus in the corresponding article (Moll and Sallnäs, 2013). The article focused a lot on universal access, specific CSCW related concepts and inclusion, while the thesis is more focused on how haptic feedback can be used for communication and how it influences the collaboration.

The second thesis study, on the same theme, was a follow-up study involving pairs of sighted and blindfolded adult students (see figure 2). In this case a between-groups experiment was carried out investigating the effects of adding audio cues to the dynamic application evaluated in the first study. A haptic condition was compared to a haptic/audio condition. The focus in this study was to get a good amount of data for statistical analysis, in order to find out if task performance was improved when audio cues were present. An important aspect considered was also how the different feedback combinations influenced the collaboration and most importantly the communication between the participants. This was the second and last study in which haptic guiding was utilized. After this study it was clear that haptic and audio functions can be used for communication in collaborative multimodal environments.

The focus in chapter 5, presenting this this study and the results thereof, is again on communication. The chapter covers all aspects of the study, while the corresponding articles cover mostly the quantitative results (Huang et al., 2012) and the qualitative results (Moll et al., 2010), respectively.

**Multimodal interface for joint task solving by sighted persons**
The third thesis study, also based on an interface developed within the scope of the thesis, concerned collaborative task solving in an abstract visual/haptic/audio environment (see...
figure 3). A within groups experiment was carried out in which pairs of sighted users solved the same task in a visual/haptic, visual/audio and visual/haptic/audio condition. The task was to use the available feedback functions to find and fill out empty spaces in a topography of stacked cubes. Again, the focus when it came to quantitative measures was on how task performance (time and errors) was affected when changing the conditions. The more qualitative focus was, again, on how changes in feedback combinations influence the collaboration, task solving strategies and most importantly the communication between the two participants. Haptic or audio guiding functions were not used in this study, but the combination of modalities was nevertheless shown to influence the collaboration and especially the communication between the participants.

The focus in chapter 6, covering this study, is very similar to that in the corresponding article (Moll et al., 2013).

1.3 Outline

This thesis is a monograph, based on results from the studies introduced in section 1.2 above and research related to those studies. The focus in the journal articles and conference proceedings produced within the scope of the studies are often somewhat different from the focus in this thesis. The thesis focuses on general themes related to communication and collaboration while the focus in many of the publications is more on the specific research areas (e.g. working with visually impaired computer users). Additionally, the thesis includes details on the studies, analysis and results not covered in the earlier publications.

Before the thesis studies are presented and elaborated on, a theoretical background including related research is provided in chapter 2. Here, the theory relevant for all or most thesis studies will be presented. The focus is to introduce key terms and related research from the main research areas studied within the thesis project. These areas include, but are not limited to, sensory modalities, multimodal interfaces and communication in multimodal interfaces.

In chapter 3 the methodologies used for analysis in the thesis studies will be subject to a theoretical discussion in order to provide a thorough methodological foundation for the work performed. The motivation for the specific studies performed will be provided in the respective study chapter.

Chapters 4-6 are the main parts of the thesis, since these will cover the thesis studies and hence the new contributions gained from my work. In chapter 4 the study on collaboration between visually impaired and sighted pupils will be discussed and the areas of collaborative learning and software support for visually impaired pupils will be briefly introduced. The follow-up study, with sighted and blindfolded adult students, is
discussed in chapter 5. In chapter 6 the study on collaborative task solving between sighted users in a visual/haptic/audio environment will be discussed. In chapters 4-6 the focus will mostly be on how haptic and audio feedback can be used to communicate information to a peer and on how changes in modality combinations affect task performance and influence collaboration.

The thesis concludes with chapter 7, where the main contributions from the thesis studies will be discussed and elaborated on further.
2 Theoretical background and related research

The studies carried out within the scope of this thesis are all focused on collaboration in multimodal virtual environments and then specifically on how the modalities supported influence our means of communicating. Research in all these fields has been carried out for decades at least when it comes to visual and audio based interfaces. During recent years research has also been carried out investigating the effects of using haptic feedback and how different combinations of the visual, haptic and audio modalities affect task performance in single user virtual environments. However, aspects such as how feedback combinations affect task performance and influence communication between users in collaborative virtual environments are still under-investigated.

In this chapter basic theories and earlier studies on collaborative virtual environments and communication in particular will be introduced and discussed. The theories and studies discussed here should be seen as a basis for understanding the thesis studies introduced later. The concepts brought up here will be more thoroughly investigated and discussed in relation to the thesis studies in their respective chapters 4-6.

2.1 Working with multimodal interfaces

One of the main focuses in this thesis is on multimodal interfaces and how we interact with them. The term modality could be defined in a number of different ways depending on the research area considered. In this thesis the term will be used in relation to the field of human-computer interaction, meaning that it will be viewed as either a sense through which we receive output from the computer or as a general classification of input equipment for the computer.

A multimodal computer interface is an interface that supports more than one input and/or output modality. Hence, one can use more than one sense when interacting with this type of interface. The research presented in this section will show that multimodal interfaces are to be preferred against unimodal ones (interfaces based on only one sense) in a number of respects. Traditionally, multimodal interfaces have been based on vision and audio but nowadays interfaces accepting speech, gestures and even haptics are developed within a constantly growing range of application areas.

Even though speech and gesture interfaces are relevant for the research questions on which this thesis is built the focus in this thesis will be on multimodal environments based on combinations of visual, haptic and audio feedback. Before a discussion of the true benefits of multimodal feedback and how sensory information from different senses are integrated and intertwined, a background about the different senses and how we use them to interact with virtual environments will be provided. This section will focus on our senses and single user interfaces and section 2.2 and 2.3 will focus on collaborative environments and communication in such environments, respectively.
2.1.1 Vision, hearing and touch in real and virtual environments

A considerable amount of research has been conducted when it comes to basic properties of our senses, cognitive abilities, perceptual organization, etc. Our senses have different strengths and weaknesses and several researchers have concluded that we, on a continuous basis, weigh the signals provided from each sense prior to forming a final perceptual impression (Baddeley, 1992; Wickens, 2002; Lederman and Klatzky, 2009) on which we react. How these weights are assigned is outside the scope of this thesis\(^2\), but an important factor considered in this section is the degree to which a certain sense is dominating. In most cases vision is the dominating sense, and hence gets a higher weight in the final percept, but for very specific tasks touch or audio can also be the dominating sense.

Here, basic information about the different senses and their use in virtual environments will be provided. Since haptic feedback is the newest and most complex feedback type when it comes to both hardware and software support, the focus will be placed on haptics and our sense of touch. The sense of vision will only be treated briefly, as a point of comparison to the other senses, since most research on vision and visual/graphical interfaces will be considered common knowledge here.

2.1.1.1 Vision and visual interfaces
Examples of tasks for which vision is dominating, and hence best suited, are recognition of patterns, discrimination of size and shape, detection of movement and gaining an overview of a complex scene (Proctor and Zandt, 1994). When exploring real and virtual environments our sense of vision is of utmost importance and therefore it is imperative that the perceptual abilities of that sense is taken into account when developing user interfaces no matter what the application area is. When exploring a complex real or virtual scene we are fast at identifying and classifying objects since we can immediately identify height relations and see the overall shape. When it comes to recognition of entire scenes we can visually explore several objects at the same time and easily see the objects’ relative positions to one another. The abilities to give a quick overview and to provide precise spatial information are maybe the most important aspects of vision to consider when developing interfaces for sighted people, especially if the scene is complex.

A lot of research has been devoted to how one should organize the elements of an environment (regardless of whether it is real or virtual) to take into account the visual sense’s ability for perceptual organization. It is well-known that related objects are grouped together by means of the gestalt principles (Proctor and Zandt, 1994) and that objects deviating from the pattern, by motion, orientation, shape, etc., are very easily spotted. It has even been shown that we perceive these deviating objects pre-attentively, without being consciously aware of it (Meng and Wang, 2009). Taking these aspects into account is important when developing user interfaces, especially in cases where object detection is imperative.

Even though the sense of vision is usually superior it is often hard to avoid some amount of visual clutter in interfaces, especially in situations in which huge amounts of information need to be presented like e.g. in scientific visualizations (Newcomb and\(^2\) The interested reader can read (Wickens, 2002) or related works by Wickens or Baddeley on multiple resource theory and performance prediction.

---

\(^2\) The interested reader can read (Wickens, 2002) or related works by Wickens or Baddeley on multiple resource theory and performance prediction.
Harding, 2006). In these situations, where a relevant focus is hard to achieve or when information is inherently occluded, the strengths of the visual sense may not be enough for effective interaction. Thus, this is one of the situations in which support for other senses than vision would be beneficial and maybe even necessary. The following sections in this chapter will mostly focus on interfaces for sighted users. Interfaces for visually impaired computer users, in which support for audio and/or haptic feedback is necessary, will be given special attention later on in chapters 4 and 5 where the thesis studies involving this special target group will be presented.

2.1.1.2 Our sense of touch
As stated earlier, vision is almost always the dominant sense, but when it comes to discriminating between and categorizing material properties like textures the sense of touch is the most efficient sense (Lederman and Klatzky, 2009). It has also been shown that the sense of touch can be used for exploring the shape and size of objects even though the resulting impression will be much more accurate if visual cues are given as well (Proctor and Zandt, 1994). Moreover, the sense of touch is better at detecting temporally distributed stimuli than the sense of vision is (Lederman and Klatzky, 2009). Temporally distributed events are events that are somehow time dependent. One typical example could be the illustration of wind flow, shown by arrows, on a weather map. Even though most people experience no problem in interpreting such a map we get a much better appreciation of the flow by actually feeling the same information. Another example, relevant for haptic computer interfaces (see section 2.1.1.3), is the one of direct manipulation of objects. Even though it is possible to move graphical objects by a 2D mouse or maybe even a 3D mouse, doing the same thing utilizing the sense of touch would give better precision and control, due to improved hand-eye coordination (Hale and Stanney, 2004).

Moving to the basics of our sense of touch, there are two ways by which humans gather information from the touch modality; cutaneous and kinesthetic perception (Lederman and Klatzky, 2009). The cutaneous part is based on mechanoreceptors and thermoreceptors in the skin and is responsible for registering and feeding through information about e.g. pressure, pain, vibration and temperature. Our sensory capabilities, when it comes to cutaneous, or tactile, perception, are not the same on all parts of our body due to the density of the receptors being different at different parts of the skin (Johansson and Vallbo, 1983). Our capabilities are most accurate on the finger tips, which have the highest density of skin receptors, where we for example can feel and discriminate between points down to 1 mm apart.

The kinesthetic part of the touch modality is responsible for providing information mostly about spatial features, like the layout of scenes, distances to or between objects and stimuli regarding position of the body. The kinesthetic sense is based on receptors in muscles and joints, providing information on the continuous change in body posture and muscle contractions. In most research the cutaneous and kinesthetic parts are treated separately and this is relevant from a theoretical perspective. Appelle (1991), however, stress the fact that we, under normal conditions, use both the cutaneous and kinesthetic parts when exploring objects. Usually, we cannot determine how object properties have been identified but we subconsciously form a unified mental image from the information gained from both the cutaneous and the kinesthetic systems (Lundin-Palmerius, 2007).
There are two ways by which we explore the world with our sense of touch; active and passive touch (Loomis and Lederman, 1986). We utilize active touch when we explore the world around us actively by directly touching and manipulating objects. Passive touch is utilized when we let objects touch us while keeping our body still. Gibson (1962) conducted an experiment on object recognition rate as a function of the degree of active touch (from purely passive touch to free hand exploration), which serves as a good example of the distinction between passive and active touch. He presented cookie cutter shapes in different ways to the subjects, who were asked to identify the shapes. Recognition was poorest (only 29%) when the respective shapes were pressed into the participants’ palms. In this case the touch was passive, since the participants were touched by the shapes and thus did not do anything to actively explore them. When the participants were able to explore the shapes freely (through active touch), on the other hand, the recognition rate was 95%. Appelle (1991) gives several pointers to other experiments yielding similar results. Gibson’s results highlight the virtues of being able to explore objects actively in comparison to being subject to passive touch. In this thesis the focus, from now on, will be on active touch.

Haptic perception is a combination of tactile and kinesthetic perception; with the constraint that it relates to active touch (i.e. you explore the world actively) (Loomis and Lederman, 1986). We seldom use tactile or kinesthetic perception alone to perceive and explore the world, but the haptic combination of tactile and kinesthetic perception. The same holds true for passive and active touch (Appelle, 1991).

While Gibson (1962) showed that object recognition is high when active exploration of objects is possible, Lederman and Klatzky (2004) and Klatzky and Lederman (1999) showed the opposite in situations where the active touch is constrained by e.g. finger sheathes or rigid probes. It is obvious that object recognition based solely on touch, in situation where the active touch is constrained, often is slow and error prone and this concerns both recognition in 2D and 3D. While detection of edges and judgment of orientation proved to be accurate when active exploration was constrained (Lederman and Klatzky, 2004) recognition of entire objects was a lot more cumbersome. However, when discriminating between textures the impairment is moderate and when touching an object through a finger sheath no effect was seen on vibratory thresholds (Klatzky and Lederman, 1999). The difference in quality of object recognition between free-hand exploration and exploration through a probe can in large be explained by our means of utilizing the sense of touch. Free hand exploration can roughly be categorized into six types (Lederman and Klatzky, 1987; Lundin-Palmerius, 2007). *Lateral motion* is utilized when moving e.g. a finger over a surface to judge the texture. As shown by Klatzky and Lederman (1999) the difference between free-hand exploration and exploration through a probe is not large in this case. *Pressure* is utilized when pressing against an object to feel hardness. *Static contact* with an object is utilized to sense temperature. *Unsupported holding*, by letting an object rest in the hand, is utilized for judging object weight. *Enclosure* is utilized when closing the hand around an object in order to judge the global shape. *Contour following* is utilized when following the object boundaries to get a feeling of the shape. When considering these exploration procedures it is easy to conclude that it is hard to feel temperature through static contact when using a probe and that enclosure is a procedure which cannot be performed by a probe. Thus, when judging the shape of an object through a probe one is bound to contour following which is slow in this case since
you only have one point of interaction in comparison to several fingers used in free-hand exploration. Additionally, you will never get the overall picture of the object as you do when enclosing the entire object in the hand. This is the main reason why identification of complex objects in haptic virtual environments is usually both slow and error prone.

Many studies have been performed focusing on the use of haptic feedback (feedback provided to the sense of touch) in virtual environments and how we can use our knowledge about the touch modality when designing virtual environments. Hale and Stanney (2004) performed a meta study investigating how one can derive haptic design guidelines from basic cognitive and perceptual abilities of our sense of touch. They go through the different cutaneous and kinesthetic receptors in detail and provide threshold values on vibratory signals, stimuli separations and forces. Studies have also been undertaken to investigate which types of parameters we use to discriminate between haptic stimuli (MacLean and Enriquez, 2003). Results indicate that the vibration frequency of the haptic stimuli dominates other haptic attributes and that the waveform used seems to be the second important attribute when judging the character of haptic stimuli. To enable discrimination between haptic stimuli on other grounds than vibration frequency the range of frequencies used needs to be small (MacLean and Enriquez, 2003). Hunter and Best (2006) also studied the abilities of our haptic system and concluded that it had both memory and recognition capabilities and that these were comparable to the visual capabilities.

When it comes to vision a lot of research has been conducted on how we perceive elements and groups of elements, how we perceive similarity and so forth. During recent years the same has been done for our sense of touch. Chang et al. (2007) conducted experiments to show that we group elements, according to the Gestalt principles of similarity and proximity, in the same way for vision and touch. They showed that similarity was judged by using color and texture respectively and that the position of elements was used to judge proximity. Studies have also been conducted on how we categorize common objects by the sense of touch (Haag, 2011). In Haag’s study the participants should name and sort toy animals based on visual only or haptic only exploration, respectively. The results clearly showed that identifying the figures in order to name them correctly was significantly harder in the haptic only condition. It was also made clear that size estimation and categorization of animal types was significantly harder in the haptic only condition. Again much of the results could be explained by the limitations in haptic exploration (Haag, 2011). Results like these and research on related topics have implications when it comes to the design of haptic interfaces in which groupings or categorization of different elements are important.

2.1.1.3 Computer haptics

Computer haptics is a rather modern research field which is “concerned with generating and rendering haptic stimuli to the human user, just as computer graphics deals with generating and rendering visual images” (Srinivasan and Basdogan, 1997). Computer interfaces based on haptics are interfaces that give feedback to the touch modality, by generating forces as the user interacts with virtual objects. Thus, these kinds of interfaces provide force feedback to the user by means of different kinds of specialized hardware. Haptic interfaces could be defined as “being concerned with the association of gesture to touch and kinesthesia to provide for communication between the humans and machines”
(Hayward et al., 2004). To be able to develop a haptic interface two components are needed; a hardware device for interaction and software to support communication with the device. In this section the hardware will be discussed. Different software will be brought up to discussion when the specific thesis studies are described in later chapters.

Device types and hardware properties

There are three main types of haptic equipment that can be used with haptic interfaces – body based equipment, table based equipment and tactile displays. Body based equipment, like exoskeleton gloves (see figure 4), is worn by the user while table based equipment (figure 5) is placed on the table or the ground. Tactile displays, like tactors (displaying vibratory patterns on the body through e.g. small transducers), display forces on a passive user. The focus in this thesis will from now on be on table based equipment.

An important aspect that all haptic feedback devices share, in one way or the other, are that they enable programmable mechanical properties that provide for a bidirectional exchange of energy and information between user and computer (Hayward et al., 2004). Thus, we can both control the environment and get feedback about its status from the same mechanical input device.

Apart from the categorization just described, there are some other properties that are important to consider when discussing haptic feedback equipment (Lundin-Palmerius, 2007):

- **Impedance vs. admittance**
  Devices based on impedance control measure the position, specified by the user, of an end-effector (e.g. the pen of the Phantom device shown in figure 5) and display a contact force through the end-effector. Devices based on admittance control instead measure force applied, from the user, to the end-effector whereby new positions are calculated for the device. Thus, in the first case the user has control over the position of the end-effector whereas the device has complete control of the end-effector’s position in the second case. One consequence is that admittance controlled devices often have superior stability since they can lock up the position of the end-effector completely to prevent the user from penetrating objects. With impedance controlled devices the user can penetrate objects if

---

maximum forces are exceeded. Most devices in use today are based on impedance control and the focus in this thesis will be on those devices.

- **Degrees of Freedom (DOF)**
  DOF denotes the number of independent dimensions which the device can read and display forces in, respectively. For example, a 6 DOF input and 3 DOF output device can read position and rotation in 3D, but only display forces along the three Euclidian axes. These devices are the most commonly used. For a 6 DOF output device not only force but also torque can be displayed. 1 DOF output devices can only display vibration or forces in one single direction.

Other important aspects to consider when choosing haptic feedback equipment are the resolution and number of interaction points. Exo-skeleton devices, like the CyberGrasp shown in figure 4, usually have several interaction points (points that can display forces in response to user interaction), but most devices in use today have only one point of interaction. One of these, which has been used in all thesis studies and most other studies referenced in this thesis, is the PHANTOM.

*The PHANTOM haptic feedback device*
In the studies presented later on in this thesis the table based PHANTOM haptic devices, developed by Sensable Technologies, have been used exclusively. The PHANTOM used in this thesis is a one point interaction 3 DOF output device consisting of a pen-like stylus attached to robot arms (example shown in figure 5 below). It comes in a variety of forms, Desktop and Omni being the most widely used in today’s research on haptic interfaces. When the graphical representation of the pen’s tip “ touched” a virtual object forces are generated which give the user the impression of actually physically interacting with the object. With the 3 DOF PHANTOM devices the user gets force feedback from touching objects in a three-dimensional environment. An alternative to using the pen-like stylus is to use a thimble attached directly to the robot arms. The thimble provides the user with a more “direct” possibility of interacting with virtual objects. Lately, 6 DOF output versions of the device have also been developed, with which the user can also feel torque.

There are different sorts of PHANTOMs with varying qualities and resolutions. One major limitation with this kind of haptic equipment is that you only have one point of interaction and thus it is hard to get a good understanding of complex objects. One point of interaction gives considerably less information than a continuous hand movement over an object (Kirkpatrick, 1999). Lederman and Klatzky’s (2004) results on object recognition through a rigid probe referenced earlier (essentially the same situation as when using a PHANTOM in a virtual environment) also show that object recognition by touch is slower when not interacting with the object directly by free hand exploration. This is much due to the lack of rich cues gained from various skin receptors during direct exploration as well as a higher cognitive memory load especially when exploring complex objects and scenes (Lederman and Klatzky, 2004). Another disadvantage is that the PHANTOMs have a relatively small work area (16x12x12 cm). In spite of these drawbacks, PHANTOM devices have been successfully used in many of the studies referenced later on in this theoretical chapter as well as in the thesis studies.
Programming haptic interfaces

During haptic interaction the computer has to go through a series of steps which have to be looped through continuously. First, the position of the proxy (the representation of the user within the environment) in the virtual environment needs to be read. Second, a collision detection needs to be performed in order to find out if the proxy has gotten in touch with a virtual object in the scene. Third, depending on whether the proxy is in contact with an object or not, contact forces need to be calculated. The force vector fed back to the user is typically dependent on the penetration depth in the virtual object and the stiffness of it, and is typically based on the well-known spring law \( F = k \times x \), where \( k \) is the material stiffness and \( x \) the object penetration depth. Even if the proxy is not in contact with an object it may still be subject to a force field or maybe a magnetic source. In the last step of the loop, the calculated forces need to be fed back to the user so he/she can feel them through for example a PHANTOM device. Each of these four steps needs to be calculated at least at a rate of \( 1 \) kHz (1000 times/second) in order for the haptic feeling to be robust enough. This update rate of course poses a constraint on the calculation performed in each loop. There has therefore been a lot of research trying to link software and hardware demands to the human perceptual capabilities, one example being the effects of decreased surface stiffness (a lower stiffness needs a lower resolution to be maintained) on perceived surface quality and discrimination of size (O’Malley and Goldfarb, 2002).

One of the major challenges regarding computer haptics concerns the simulation of deformable objects. One research area in which deformable object behavior in haptic simulators can make a considerable contribution is medicine – a scenario where surgeons could plan strategies and even do virtual operations in a realistic environment based on both vision and touch, before the real operations, could prove to be invaluable. The problem of simulating deformable object behavior has not come to a complete solution yet, much due to the constraint that an update frequency of at least \( 1 \) kHz needs to be maintained. The calculations needed for a realistic feeling of a deformable behavior are complex and extremely resource demanding and up until today only approximations have
been developed (Balanuik, 2006). The results are that, up until today, one needs to
neglect important properties of deformable tissues when developing simulators based on
these kinds of objects.

**Haptic icons – an example of haptic feedback in computer interfaces**

Before moving on to the sense of hearing and auditory displays, an example of the use of
haptic feedback in computer interfaces – haptic icons – will conclude this section. Lately,
a lot of research has been devoted to the study of hapticons/tactons or haptic/tactile icons
(MacLean and Enriquez, 2003; Enriquez and MacLean, 2003; Brewster and Brown,
2004; Brown et al., 2005; Hoggan et al., 2009). Haptic icons can be defined as “brief
programmed forces applied to a user through a haptic interface, with the role of
communicating a simple idea in manner similar to visual or auditory icons” (Enriquez
and MacLean, 2003). Examples of such icons could be a continuous sinusoidal wave
representing a save procedure or a short vibration representing the closing action in a text
editor. Although most studies on haptic icons consider their use in mobile phones, they
could be used to convey information to the touch modality in any kind of interface.

Common attributes to use for haptic icons are frequency, wavelength (sine, saw tooth,
etc.) and force amplitude and as stated earlier one has to take human perceptual abilities
and threshold values into consideration when choosing parameters (MacLean and
Enriquez, 2003). Although success in recognition has been reported in this and related
studies, Brown et al. (2005) have shown that you can get even more pronounced results
when using more complex waveforms and rhythmic patterns. They also showed that
haptic icons can be successfully classified based on perceived roughness and that
different combinations of roughness/rhythm in the icons can be used to successfully
distinguish between different messages in a computer interface. Brewster and Brown
(2004) also differentiate between different types of tactons; compound, hierarchical and
transformational. This taxonomy is the same that McGookin and Brewster (2004) used
for earcons (see the discussion about earcons in the next section) and should be easily
transferrable to hapticons as well. A more thorough discussion about the communicative
aspects of hapticons will be provided in section 2.3.2 about mediated touch.

### 2.1.1.4 Our sense of hearing and auditory design

Hearing is the best sense when it comes to detecting temporality or temporal changes
(Proctor and Zandt, 1994; Lederman and Klatzky, 2009) and thus is the superior sense in
situations where one needs to judge time varying properties like velocity. If we refer back
to the example of wind flow given in the beginning of section 2.1.1.2, hearing the wind
flow would give us an even better appreciation of the flow than if we only felt it. More
generally we are better at judging speed and acceleration with the sense of hearing than
with the other senses. Since our hearing is also highly sensitive to stimuli it is also often
used to direct attention to specific events or areas of importance through e.g. warning
signals or alert messages.

Information is provided to the sense of hearing through pressure waves in the air and
auditory perception is all about making sense of these waves. The hearing system is
highly complex and filtering of information and a number of interactions between short
and long term memory are performed before the meaning of a sound is deduced. There
are a number of properties, related to the waves, that are linked together to form a final
percept; frequency, pressure level, wave shape and time interval (Sethares, 2007). The frequency concerns our perception of pitch, which allows a kind of ordering of the sound on a musical scale. Note that pitch is a completely subjective measure. The pressure level corresponds to our perception of loudness, which enables ordering on a scale from quiet to loud. The wave shape corresponds to our perception of timbre, which is used to deduce the type of sound (voice, string, organ, etc.). Last, the time interval of sounds corresponds to our perception of duration which enables us to sort the sound on a scale from short to long. The most important thing to understand here is that a sound is not really perceived as having a specific pitch, duration, etc. but rather as something concrete that has a meaning to us (Sethares, 2007), like a car accelerating or a phone ringing. It is also clear that the human auditory system is very capable of masking out specific auditory events which are by some means unusual, depending on the current context (Chardin, 2008).

This, of course, has implications for auditory interface design especially when warning and confirmation messages are concerned.

The principle of perceptual organization (refer to the gestalt principles) which applies to vision and touch also applies to hearing (Proctor and Zandt, 1994; Sethares, 2007; Chardin, 2008). Temporal proximity is measured by the time interval between stimuli – sounds coming in close succession tend to be grouped together. Similarity is measured by the pitch or frequency of tones. Even though our sense of hearing can also be used to judge spatial proximity, hearing is not primarily a spatial sense. We can detect the position of signals, roughly, in 3D and the well-known Doppler-effect can be used to detect changes in position due to variations in pitch.

When it comes to virtual environments a lot of research has been conducted on how to sonify whole environments, create interaction sounds or develop sound icons (Gaver, 1993; Brewster, 1994; Winberg and Bowers, 2004; McGookin and Brewster, 2004; Rassmus-Gröhn, 2008; Winberg, 2008; Hoggan et al., 2009). In the work performed by Winberg (2008), on accessible interfaces for blind computer users, the following definition of sonification is used: “The transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation”. In a sonification whole environments are represented by audio and the idea is often to represent different kinds of data by acoustic signals. Earcons have been defined by Brewster (1994) as “abstract, synthetic tones that can be used in structured combinations to create auditory messages”. The difference between sound icons and earcons is that sound icons are specifically designed to be intuitive (close relation to the aspect sonified) while earcons are not. A formal definition of sound icons has been provided by Gaver (1993); “everyday sounds mapped to computer events by analogy with everyday sound-producing events”. Thus, sound icons are meant to resemble sounds experienced in real life. One example of a sound icon could be the paper rustle heard when the trash can on the computer desktop is emptied and one example of an earcon from the same domain could be the abstract sound heard when an email is received.

According to Rassmus-Gröhn (2008) sound icons are often easier to learn than earcons since sound icons can give a direct association to real-life events if designed appropriately. A higher cognitive load is usually associated with remembering and interpreting earcons. When there are no real-life equivalents to an action in the real world, however, a natural sound for a sound icon may not be possible to find (Brewster, 1994).
McGookin and Brewster (2004) have studied earcons in depth. They differentiate between four different types of earcons. One-element earcons are earcons which are based on a single property, like pitch or timbre. Abstract and immediate interaction sounds, providing feedback on actions performed in an interface, are usually based on this type of earcon. Compound earcons are one-element earcons grouped together to form a kind of message. One example that is given is that the sound for “open” followed by the sound for “file” forms a compound earcon providing the message “open file”. Hierarchical earcons are earcons that are ordered in a kind of hierarchical tree structure. The earcons on a particular level in the tree “inherit” the meaning from all parent nodes. The leaves of the tree constitute earcons on the most detailed level. The tree structure and the hierarchical dependencies construct a kind of grammar based on earcons (McGookin and Brewster, 2004). One specific advantage of this is that you only have to learn the rules for construction of earcons and not each individual earcon. Last, transformational earcons are also related to grammar but the parameters which constitute the earcon (pitch, timbre, duration,...) can be altered to change the meaning of the earcon. In the thesis studies both sonifications and earcons have been used to convey information in collaborative environments. The details of the implementations will be provided in later chapters.

As an example of the use of sonifications a study by Winberg and Bowers (2004) will be used. They developed a Tower of Hanoi game (in which disks should be moved between three pegs) based solely on auditory feedback, to be used by visually impaired computer users. They developed two versions of the interface with three and four disks respectively and each disk was associated with its own sound differing in pitch and timbre from the others. Stereo-panning was used to convey information about on which peg a particular disk was placed and the duration (length of the sound) was used to convey information about the level of the respective disk. A regular mouse was used to move a disk from one peg to another. The sound for each disk, altered to give information about placement as described above, was looped until the game was solved, to provide a continuous “view” of the state of the game. An evaluation of the game showed that it was possible to solve the game both with a visual interface (the traditional way of solving the game) and with a completely sonified environment with no visual cues given. With their study Winberg and Bowers (2004) clearly show that auditory feedback can be used for successful task completion in a virtual environment and they also illustrate how the perceptual dimensions of sound may be utilized.

Research has also been conducted on how to find objects in virtual environments by means of audio. Crommentuijn and Winberg (2006) compared different means of finding objects in a virtual space. They found that an implementation where the sound source was positioned at the location of the virtual object was the solution often generating the shortest object localization time. This solution was found to be intuitive and it allowed for fast tracking of object positions. This implementation of sound, for locating objects in 3D virtual environments, has been used successfully in several studies and is called “ears in hand” (Magnusson et al., 2006; Murphy et al., 2012). The metaphor focuses on the fact that the user experiences the sound environment as if his or her ears were bound to the proxy moving around in the environment. This metaphor has been used in some of the thesis studies discussed in later chapters.
2.1.2 Cross-modality relations

Traditionally, the general view of our sensory systems was that they worked in parallel and independently without any interaction. During the most recent decades, however, this view has changed (Shimojo and Shams, 2001). Nowadays it is acknowledged that there is a lot of interaction and inter-dependencies between our different senses and this must of course be taken into account when developing and studying interaction in multimodal interfaces. The interaction between sensory modalities is complex and even today a lot of questions about why interaction takes place and how it is done remain unanswered. The term cross-modal, not to be confused with the term multimodal, refers to this interaction between the senses. Cross-modal interaction is the process of coordinating information from different sensory channels into a final percept (Driver and Spence, 1998).

During recent decades a large amount of research on cross-modality relations has been conducted, aiming at understanding how the different senses interact with one another and how one can use these interactions to enhance the experience and work efficiently in both real and virtual environments. Many studies have been carried out e.g. investigating how stimuli sent to one sense can trigger attention in another (Driver and Spence, 1998; Gray et al., 2002; Gray et al., 2009; Jones et al., 2007) and how integration of signals from different senses work (Warren and Rossano, 1991; Sathian, 2005; Ernst and Banks, 2002). Even though research on attention in various situations is important that topic will not be covered in more depth in this thesis due to its small relevance to the research carried out during the studies presented in chapters 4-6.

One issue of especial interest when considering visually impaired people, as was done in a few of the thesis studies, is that of plasticity. The term plasticity in this respect refers to the brain’s ability to change its use and delegation of resources in response to events such as injuries or learning of new material. For example, if vision is lost during the early years of life the resources in the brain related to vision can be used to boost another sense and as a result congenitally blind persons often have a more sensitive sense of touch (Shimojo and Shams, 2001). This has to be taken into account when using blindfolded users instead of blind users in studies. In a recent study by Collignon et al. (2009) it was also shown that blind subjects’ spatial processing of sounds was much enhanced compared to that of normally sighted persons. This is due to the fact that the resources in the brain originally used by the visual sense are now used by the fully working sense of hearing. Another finding that shed light on the plasticity of the brain and visual dominance is that early vision (i.e. blindness occurring after some years with vision) greatly influences tactile perception in the blind (Röder et al., 2004). Every time a blind person, who has been sighted, uses the sense of touch, he/she will transform the sensation into a visual mental image before deciding where a tactile stimulus came from and what it means. On the other hand, persons who are congenitally blind (borne blind) will never do this transformation and will therefore often react faster to haptic stimuli. More about haptics in relation to visual impairment will be provided in chapter 4 and 5, where the thesis studies on collaboration between sighted users and users that cannot see are presented. In a large number of studies it has also been proven that areas in the brain related to the processing of purely visual information are also activated when processing haptic input (Sathian, 2005), showing that vision and touch are interrelated by nature.
Our means of interacting multimodally and the cross-modal relations that exist between the senses also have more subtle implications for interface design. When we interact with multimodal interfaces our brain always tries to fuse the information conveyed from the different modalities (often vision and touch and/or audio) to construct a unity. Hoggan et al. (2008) explored this in an experiment where users paired the haptic/audio feeling of different touch screen buttons with their graphical appearance. The overall goal was to find out if it was possible to find congruence between the look, feel and sound of different buttons. The results indicate that users have a clear sense of how different kinds of buttons should feel and sound. More interestingly, a follow-up questionnaire on user satisfaction showed that the combinations judged to be most congruent in the experiment were the ones rated as most appealing. McGee et al. (2002) conducted an experiment on a similar theme when it came to discrimination of textures. Among other things it was shown that audio cues that were congruent with the roughness of a texture clearly affected the perception of the texture roughness. Thus, it is clear that the audio adds to our perception of texture and this should be taken into account when constructing interfaces where discrimination among several different textures is important (McGee et al., 2002). In a later study Hoggan et al. (2009) investigated which type of parameters in audio and tactile icons respectively should be used to convey different types of messages. In an experiment participants should map different parameters of both audio and haptic icons to different kinds of messages. It was clearly shown that it is possible to design cross-modal icons (in this case audio/tactile) which do have an agreed-upon meaning. Thus, when constructing a multimodal user interface the rate of congruence between look, feel and sound as well as the specific parameters used to convey messages may very well affect the overall acceptance, performance and user enjoyment.

Another important aspect of multimodal interface design, related to cross-modality relations, is that illusions easily occur due to the dominance of the visual sense. Illusions may result in poor interface design if they are not spotted in the design phase. One illustrating example of a perceptual illusion was constructed by Srinivasan et al. (1996). They compared the perception of stiffness of two virtual springs. The task was to decide which one of two virtual springs was the stiffest (one was always stiffer than the other), by suppressing them by using a 3 DOF haptic feedback device. While compressing the springs the subjects were also told to look at a visual representation of the compression. At some trials the visual feedback and the tactile feedback corresponded in that visual displacement was the same as the tactile displacement felt through the touch feedback device. The participants gave the correct answer about the stiffness on the vast majority of these trials. However, as the visual representation deviated from the haptic one (e.g. the visual compression of one spring was much bigger than in reality) discrimination of the stiffest spring was much harder. When the visual compression of the stiffest spring in reality corresponded to the less stiff one the responses were wrong on the vast majority of the trials, indicating that the subjects disregarded the information given by the hand position and only cared about what they saw. Thus, by graphically showing a compression that is much larger than the one actually felt, one can be given the impression of touching a much softer surface. The phenomenon discovered by Srinivasan et al. (1996) is especially relevant when it comes to interface design for visually impaired computer users. Even though the example given above is constructed it shows how easy
it is to introduce haptic design artifacts which may not be spotted during the design phase. One consequence of this is that a sighted designer can, without noticing it, create environments including elements that are hard to understand for blind users. More about the challenges of designing for visually impaired computer users will be provided in chapter 4. Apart from the illusion described by Srinivasan et al. (1996) several other visual-haptic illusions also exist. One example is the size-weight illusion – if two objects with the same physical weight have different dimensions, but look as if they are of the same material, we perceive the smaller one as being heavier than the larger (Koseleff, 1957). Researchers have also found effects of shape and material on perceived differences in heaviness (Ellis and Lederman, 1999).

One of the most widely known audio-visual illusions, which also has implications for interface design, is the McGurk effect (Ali and Marsden, 2003). The effect is dependent on temporally aligned visual and auditory stimuli being in conflict with one another. The perceived sound could differ from the information provided to both the visual and the auditory sense. The McGurk effect is mostly known for its application in audio-visual speech synthesis and clearly shows that the exact same sound can be perceived in different ways depending on the visual stimuli provided.

### 2.1.3 Benefits of multimodal input

There are many reasons to why multimodal feedback is preferable to unimodal (one sense) feedback. As already discussed all senses have their strengths and weaknesses and are used for different purposes in virtual environments. Since our sense of hearing is especially good at detecting temporal changes and has a fast reaction time, audio feedback has often proven to be beneficial for providing different kinds of warning messages and feedback on actions taken in both real and virtual environments (Cao et al., 2010; Sethares, 2007). On the other hand, our sense of touch has been shown to be highly suitable for judging surface details like textures and for manipulation of virtual objects and hence could be of much value for task performance in interfaces where e.g. object manipulation is important (Sallnäs, 2004; Gupta et al., 1997). More generally, using more senses than vision when working in both real and virtual environments has been shown to reduce workload on the visual sense and increase redundancy and confidence in the results achieved. Using more than one sense in the interaction with a computer interface can also give a triangulation in that interfaces can be explored from more angles. This in turn gives a richer experience and is probably one of the main reasons why task performance is usually increased when support for more senses is added to computer interfaces. This effect could be expected to be pronounced if the strengths and weaknesses of the respective sense are considered during the design. In this section the virtues of multimodal input introduced above will be further discussed, in relation to single-user interfaces. In section 2.2 we will then move on to collaborative multimodal interfaces which are most central to the thesis studies.

A large amount of research on multimodality focuses on decreasing the load on the visual sense since the ability to distribute signals to different modalities is one of the virtues of developing multimodal interfaces (see e.g. Burke et al., 2006). One clear benefit of multimodal input, related to decreasing visual load, is inherent in the research
on haptic and audio icons discussed earlier (Brewster and Brown, 2004; Brown et al., 2005; Hoggan et al., 2009). All these studies show that it is possible to create discriminable hapticons and earcons that convey simple messages like confirmations, warnings and different kinds of alerts. These messages could be both simple and compound and even convey hierarchical information. When working in interfaces which place high demands on the visual sense (more and more common today), being able to present meaningful information to other, not overloaded, senses often proves to be of importance. Two common and popular examples of computer interfaces placing high demands on the visual sense are air-traffic control and power plant interfaces. Both of these present an overwhelming amount of information to the visual sense and without audio icons/earcons, and in some cases haptic feedback, presenting warnings and/or alerts it would be impossible to direct attention to the appropriate interface element at all times.

Another advantage of working in a multimodal interface, compared to working in a visual-only one, is that the addition of support for especially the touch modality can enhance the experience of the interaction by conveying additional information not easily conveyed through vision (Hamza-Lup and Adams, 2008; Davies et al., 2009; Vines et al., 2009; Wang and Quek, 2010). One example is scientific visualizations where support for the touch modality can make it possible to convey abstract information. For example, Vines et al. (2009) developed a visual/haptic system for real-time display of mixing fluids, providing a completely new means of experiencing fluid dynamics. Hamza-Lup and Adams (2008) developed an e-learning system for visualizing concepts of pressure and hydraulics and Davies et al. (2009) developed a system for learning of molecule interactions. Especially when it comes to Davies et al’s (2009) study the potential of the added touch modality, conveying information about forces and dynamic interactions not easily grasped through reading a textbook, should be clear. Wang and Quek (2010) carried out a study in which different kinds of affective touch were conveyed at the same time as an emotionally engaging story was told. It was clearly shown that the emotions felt by the listeners were different for those who got added touch feedback compared to those in the control group who only listened to the story. The above examples should make it clear that the sense of touch has great potential in both educational settings and when trying to visualize, or maybe conceptualize, information not easily conveyed through text or speech.

Even earcons and auditory icons have shown to be a great aid in e-learning and when conveying additional information on material presented by voice or text. In their study with an interface including written material and a human-like speaking proxy (the teacher) Alseid and Rigas (2011) investigated how learning and usability would be affected by sound cues highlighting the most important material. Interestingly enough, results on a post-test showed that auditory cues, in forms of icons and earcons, presenting additional information about the learning material (level of importance) significantly assisted the users’ learning performance. More importantly, the study shows yet again that one can use another sense than vision to successfully distribute cognitive load and that this has pronounced effects on performance.

Another means by which haptic feedback has been shown to add quality in visual environments is provided by its ability to enhance depth perception and judgment of size. The meta study performed by Hale and Stanney (2004) showed that the additional information gained from kinesthesis enables more precise judgment of distance in
particular. This in part explains why haptic feedback is beneficial in interfaces supporting direct manipulation of virtual objects. With a better sense of distance in 3D, provided by the support for the sense of touch, finding and controlling objects in space will be easier than if only a visual interface is provided. The same argument holds for the positioning of interface elements in virtual environments.

Burke et al. (2006) performed a meta-study on how haptic and audio feedback respectively affected task performance, both with regard to error rates and time to complete tasks, when added to purely graphical interfaces. The study clearly shows pronounced benefits of combining information from different sensory modalities. They considered 43 studies performed during the 1990 and 2000 decades and found pronounced effects when it came to benefits of multimodal feedback. The key elements studied were task type, number of tasks and workload and the tasks ranged from target acquisition and display of warning to navigation and communication.

The addition of audio feedback to traditionally visual interfaces was shown to have a significant positive effect when it came to reaction time and task performance scores except when it came to reduction of errors, where no effect of multimodality could be found (Burke et al., 2006). However, the effects of adding audio to a visual interface were much more pronounced for single task scenarios and in cases where the cognitive workload was moderate. For most multi-task scenarios studied and for cognitively demanding tasks the addition of audio did not have a pronounced effect on task performance scores and reaction times. Burke et al. (2006) argue that this is a result of the visual and auditory modalities partly sharing the same cognitive resources. This is in line with the multiple resource theory, stating that the benefits of working with demanding tasks increases when we can distribute workload between separate resources (Wickens, 2002).

When it came to addition of haptic/tactile feedback to a visual interface Burke et al. (2006) also found significant positive effects. As was the case when studying the effects of adding sound to a visual interface, performance scores were improved and reaction times decreased when haptic feedback was added. Furthermore, no significant effect could be seen when looking at error rates. Interestingly enough, haptic feedback was shown to give a pronounced benefit in a case where audio did not – for multi task scenarios. Even in cases where the cognitive workload was demanding the addition of haptic feedback showed significant effects. These effects were also noted for all task types studied, except navigation, where the effect was large but not significant.

During the 2000 century researchers began to investigate how the combination of the three senses vision, touch and hearing affect task performance, showing that tasks are carried out faster and with greater precision when support for all three senses are available. Interestingly enough the studies involving modality combinations of three senses are still quite few in 2013 and most of them seem to concern interaction by visually impaired. One example of such a study, showing a pronounced benefit of using trimodal feedback as opposed to unimodal feedback, was carried out by Lee and Spence (2008). They performed studies with sighted users in which they compared the effects of the different modality combinations vision, vision/haptic, vision/audio and vision/haptic/audio on performance in a dual task scenario. In this particular study the participants were driving a car in a simulator, trying to avoid collision with another car in front of them, as they performed a series of simple tasks on a touch screen representation
of a mobile phone. The haptic feedback was presented to the participants as short vibrations for button click events and the auditory feedback was presented as a bell sound heard indicating button click events. Results showed that trimodal feedback was superior to unimodal feedback both when it came to reaction times when braking to avoid collision and time to complete the tasks on the mobile touch screen. Although an improvement could be seen when adding either haptic or audio to the visual only condition no significant differences were found in those cases.

Even though studying blind or visually impaired users’ interaction in virtual environments is different from studying sighted ones the study performed by Yu and Brewster (2003) on blind users’ interaction with bar charts, tells us a lot about the benefits of adding support for more senses to a virtual environment. In that study blindfolded sighted participants should scan a bar chart either by using only haptic feedback or by using a combination of haptic and audio feedback. One of the haptic feedback devices used was the PHANTOM Desktop introduced earlier, with which the bars and workspace boundaries could be felt (a haptic mouse was also used as a comparison to the PHANTOM Desktop, but it will not be covered here). The audio feedback gave information about specific values of bars through speech, on demand, and continuous information about bar values conveyed through the length of tones played whenever the proxy got in touch with a bar. The participants were presented with a series of graphs and a couple of related questions about trends and specific values. The results showed, again, that the condition with two senses supported was significantly superior with regard to task completion times. Also, a self-evaluation, through a questionnaire, performed after the experiment showed that the participants’ perceived mental workload was significantly lower in the haptic/audio condition indicating that the participants preferred to work in the condition with two senses supported and that they felt that the tasks were performed with higher efficiency in that condition.

This concludes the introduction to human senses and multimodal interfaces. Each sense has its unique strengths which distinguishes them from the other senses. We seldom use one sense only when interacting with our surroundings – even though vision is the primary sense there are circumstances in which that sense falls short. This is why one can often benefit from using support for more senses than vision in virtual environments. Thus, it is clear that there is a potential in implementing support for more than one sense in virtual environments. The meta-study performed by Burke et al. (2006) does not only highlight this potential, but also that different modalities are best suited for different kinds of tasks. Lee and Spence (2008) and a few others show the same thing for three senses. This is important to take into consideration when designing multimodal systems. That the addition of haptic and audio feedback to single user systems can increase performance and user experience is clear, but the next section will show that this fact holds true even in collaborative virtual environments, which are central to the thesis work.

2.2 Collaborating in virtual environments

There should be no doubt, after reading the above sections, that both haptic and audio feedback and especially the combination thereof can affect user performance, attitudes and the sense of immersion in single user virtual environments. Studies on the effect of
modality combinations on various aspects have been carried out for several decades. Studies on collaboration in multimodal interfaces have also received increased attention during the last decade and even in these kinds of environments haptic and audio feedback respectively have been shown to have pronounced effects on various aspects of collaboration. There are, however, very few of these studies that directly study the influences of modality combinations on collaboration and communication in particular. This is why one of the main purposes of the thesis work has been to shed light on how the modalities supported in the interface influence collaboration and communication in collaborative virtual environments.

This section will first and foremost discuss important aspects of collaborative virtual environments and introduce studies in which the effects of haptic and audio feedback on collaboration are highlighted. Since interpretation and understanding of the thesis studies’ results requires knowledge of the CSCW field, the section will start off by introducing some CSCW specific concepts and position the thesis work in relation to these. Even though communication between users is an important aspect to consider in collaborative virtual environments, communicative functions will be discussed in its own section 2.3.

2.2.1 Important terms and definitions

When studying collaboration in any kind of interface there are some aspects which are of special importance. Three of these; awareness, common ground and deixis, which are most central to the thesis work, will be introduced here.

2.2.1.1 Awareness

Within the field of CSCW the term awareness relates to the knowledge one has of other group members’ actions. This knowledge is used to inform one’s own action in a way that makes the whole group move forward in the collaborative process. In other words, for collaboration to work each group member must provide awareness information to the peers and at the same time process awareness information provided by the peers. The most widely used definition of awareness is probably the one given by Dourish and Bellotti (1992) related to their well-known study on shared editing; “an understanding of the activities of others, which provides a context for your own activity”. Thus, to be able to successfully contribute in collaborative work a good sense of awareness is a necessity. This view on awareness is very different from the concept of shared mental models (Jonker et al., 2011). The difference between the concepts is that shared mental models relate more to structured work teams in which the members have specified roles.

The awareness concept is complex and it has been up for debate (see e.g. (Schmidt, 2002), where the wide use of the term is subject to critique). Also, there are many different types of awareness information; social awareness, action awareness and location awareness to mention a few of the more commonly used. In one of the many attempts to clarify terminology Carroll et al. (2003) discriminated between three different types of awareness information relevant (but not exclusively) for the study of collaboration in virtual environments; social awareness, action awareness and activity awareness.
Social awareness in this case, refers to the extent to which collaborating partners perceive each other’s presence in the shared environment. This awareness is needed in order to feel that one is a part of a group working together. The term social awareness also concerns knowledge about who is currently online/on site and thus available for discussions and collaboration. Action awareness concerns handling of shared objects. If a group of people, by some means, work together on manipulating the state of shared objects it is of utmost importance that knowledge or awareness of the continuous changing state of these objects is fed through to all persons involved so that coordination of joint actions can be successfully made. According to Carroll et al. (2003) activity awareness refers to the awareness of the state of the major goals and the overall situation in which specific actions are carried out. This kind of awareness is crucial to, in the long run, coordinate actions to meet original work requirements set up before the collaborative work started.

In face-to-face collaboration acquiring awareness information on all the above levels is usually no problem since everyone can relate to the same environment and can get at least a peripheral continuous understanding of the status of the joint work material. However, gaining this information through mediated technologies has often proven to be difficult (Carroll et al., 2003; Convertino et al., 2011) and much research has been devoted to designing shared interfaces for optimal awareness information support. One attempt to form a framework for awareness support in real-time groupware (based mostly on visual feedback) was made by Gutwin and Greenberg (2002). The framework states what kind of information is necessary in collaborative environments, how awareness information is usually gathered and how awareness information is typically used in collaboration for a variety of tasks. Knowledge about all these areas is necessary to design a truly usable collaborative system where the relevant information about the doings of others and the status of others’ work is fed through in a non-obtrusive but yet informative manner. A recent attempt to provide a framework for awareness systems was also developed by Gallego et al. (2011), who proposed a formal notation for different types of awareness, how these types are linked and what workspace features that support each type. Although several attempts on formalizing the work on integration of awareness information in collaborative systems have been proposed lately, the issue of awareness still poses a real challenge to designers.

2.2.1.2 Common ground
Common ground is much related to awareness in that both concern prerequisites for effective communication and collaboration. These two concepts are partly overlapping but there are important differences. While awareness is more about knowledge of what is going on at a particular moment, common ground refers to the common understanding of joint goals, shared resources and the state of the task solving process. Common ground also refers to shared history and culture, which are both important for being able to reach a shared understanding. This common understanding is built up over time and refined during conversations or task solving processes. Clark (1996) describes that “mutual belief, mutual knowledge, mutual assumption and mutual awareness” are all subcategories of common ground. More formally common ground is defined by Clark and Brennan (1991) as a state of mutual understanding among conversational participants about the topic at hand. This mutual understanding about the topic and also about shared goals and
resources is of course of utmost importance for effective communication and collaboration. If a group of persons do not share a common ground about the topic at hand it is necessary for them to achieve this common understanding before effective communication and collaboration can take place.

A lot of research has been devoted to studies on how we establish and then maintain a common ground during collaboration and conversation (Clark and Wilkes-Gibbs, 1986). Most research in this area has concerned conversation and how we use different types of acknowledgements, references to objects, etc. to continuously add to the shared frame of reference. The process of establishing common ground is called **grounding** (Clark and Brennan, 1991; Roque and Traum, 2008). Grounding is a continuous process in that it is needed to always keep the shared frame of reference up to date. For example, as soon as a new object (or more generally; referent) is being introduced during a conversation between two persons, they need to use grounding techniques to make sure that the reference is being fully understood by both (Clark and Wilkes-Gibbs, 1986). Of course, the grounding techniques used will differ depending on if the object could be seen by both persons (a simple gesture followed by a short acknowledgement may be enough) or if the object is located elsewhere (more complex object definitions and location descriptions may be required before the object can be referred to in the conversation).

Roque and Traum (2008) performed a corpus study on simulated radio communication in a virtual environment and related the findings to different types of grounding identified, which they call “evidence of understanding”. The taxonomy they use serves as a good explanation of common ground and how it is achieved and maintained. A step that was always taken in conversations was the **submit**, which is basically an addition of something new to the conversation or the start of a conversation. The **repeat back** type of grounding is a kind of repetition serving as a confirmation. This type of grounding is especially common when grounding different kinds of verbatim content (Clark and Brennan, 1991). The **re-submit** type of grounding is a kind of “negative evidence” often used as a self-repair. **Acknowledgements** were often used to give short confirmations, like “Roger” and “Ok”, that the message had been received. Given that there is no need for self-repairs, acknowledgements are often evidence that the conversational partners share a good amount of common ground. The **request repair** type of grounding indicates, by utterances like “please, repeat!”, that the message needs to be sent again. Thus, the receiver or conversational partner does not yet feel that a common ground is shared. The **move on** type of grounding provides evidence that the last step is sufficiently grounded, by one of the partners moving on to the next. According to Clark and Brennan (1991) the initiation of a new topic in a conversation reaches a similar effect. The **use** type of grounding is a kind of response that is a relevant continuation of the previous utterance. It could, e.g. be an answer to a question or the submission of requested information. Of course, this type of grounding is in itself a confirmation that an utterance has been understood. The last type of grounding used in Roque and Traum’s (2008) study is **lack of response**, which can mean different things depending on how long the pause is. Usually the lack of response can be seen as a confirmation.

One important type of grounding, which is not covered by Roque and Traum (2008) is the use of gestures, often in the form of deictic references (more about this special kind of gesture in section 2.2.1.3 below). Gestures can be anything from head nods for acknowledgements to symbolic illustrations of a procedure to follow. Several studies
have shown that the possibility to use gestures greatly simplifies the establishment of common ground and increases the overall task performance for both remote physical tasks (Kirk and Fraser, 2006; Kirk et al., 2007) and collaboration in virtual environments (Verhulsdonck, 2007). There are several types of gestures which can be used in conversations. In a meta-study on the use of gestures in virtual environments Verhulsdonck (2007) bring up iconic, metaphoric, deictic and beat/cohesive gestures. This taxonomy has not been used explicitly in the thesis studies, but is nevertheless important for the understanding of the kinds of information that non-verbal gesture communication can convey. Iconic gestures are mostly used to complement what is being said, conveying additional semantic information. An example of such a gesture is when a speaker complements the verbal description of an object by illustrating its dimensions by the distance between the thumb and the index finger. The prerequisite is that both conversational participants have a shared understanding of the specific gestures used. Metaphoric gestures are used to represent a more abstract meaning in the same way as metaphors are used to explain an aspect of a complex domain by using analogues from more known domains. Beat/cohesive gestures are used to signal elements of special importance to the things being said. Deictic references, which are used to point out specific referents, will be more thoroughly described below, since this type of gesture is most central to the thesis. In summary, gestures can be used to convey many types of non-verbal information and are thus a good means of establishing common ground among conversational partners. In light of this it is not difficult to see that the inability to use gestures could place considerable constraints on collaboration. The importance of being able to use gestures is shown in all thesis studies performed.

2.2.1.3 Deixis
A special type of grounding, used to make references, is the use of deixis and this type of grounding deserves a special section here due to its importance to the thesis studies. Deictic references are often used to refer to or indicate objects or places by using words like “this”, “that” or “the one to the right” and/or by pointing (Clark, 2003). In most cases utterances like the ones above are combined with a gesture to further ease the identification of the correct referent. It is easy to see why being able to use deixis in conversation (or collaboration in general) is highly beneficial – describing objects or other relevant referents verbally can soon become cumbersome and an insufficient description can easily lead to misconceptions that will hinder the establishment of common ground. In fact, one of the thesis studies elaborated on in the discussion section 7.4 shows clear benefits of adding support for deictic referencing during medical team meetings (Sallnäs et al., 2011). Verhulsdonck (2007) also states that deictic gestures are the most crucial type of gestures to implement support for in collaborative virtual environments.

Making a deictic reference is one means of indicating objects or referents in a conversation and this kind of referencing falls into the category which Clark (2003) denotes “directing-to”. Clark contrasts this kind of indication against “placing-for”. While “directing-to” indications, like a deictic reference, is used to direct attention towards an object, placing is used to physically place an object into the addressee’s focus of attention. Clark (2003) exemplifies placing as putting items to be purchased on the counter in a shop – the customer places the items in the clerk’s current focus of attention.
signals that he/she wants to buy them. Note that this placing action will only have an effect if the customer and clerk have a common ground about the meaning of this particular action. Both deictic referencing and placing will be discussed later on in this thesis, but the focus will be on deixis.

In conversations in real life making use of deixis is usually not a problem, as long as the referent and of course the gestures made can be seen by the conversational participants. Supporting and using deictic referencing in collaborative virtual environments often prove to be harder (Wong and Gutwin, 2010), even though the need of doing so is evident due to their potential of providing both efficiency gains and richness in the communication (Verhulsdonck, 2007). Kirk et al. (2007) studied how the communication between a worker and “a helper” during an assembly task was affected by the ability to use deictic referencing. Two versions were compared; one in which only verbal directions were possible to use and one in which both verbal referencing and gestural referencing (provided through a projection of the workers hands on the helpers workspace) could be used. It was clear that task performance, measured as time to complete the task, was increased when a combination of verbal directions and gestures could be used. Kirk et al. (2007) also showed that there were significantly less words spoken when deictic references could be provided. This held true for both the worker and the helper. In an earlier study Kirk and Fraser (2006) compared different means of conveying gestures and concluded that projection of hands was the best alternative. Ou et al. (2003) studied collaboration on physical tasks, based on an application in which it was possible to draw in a separate layer in a video stream. The pen could also be used to control cameras mounted on the respective sites. Initial user studies showed that task performance was significantly faster in the system enabling remote gesturing than in a simpler system in which no remote drawing could be made. Wong and Gutwin (2010) also investigated the accuracy of deictic referencing in virtual environments, acknowledging the importance of being able to use these kinds of gestures in collaborative virtual environments. In one experiment they studied the accuracy in determining the direction of pointing under several different conditions. They found that accuracy was almost always better in the real world even though the differences were not very big. They also found that pointing was more accurate the further away the referred object was.

### 2.2.2 Awareness and common ground in this thesis

Since the related terms awareness and common ground are highly relevant in this thesis their relation to the thesis studies will be explained here. This discussion will be deepened in later chapters about the studies.

As stated earlier the term awareness has been up for debate (Schmidt, 2002) and therefore it is important to state that the definition of awareness used in the thesis is that given by Dourish and Bellotti (1992); “an understanding of the activities of others, which provides a context for your own activity”. Traditionally, collaborative studies investigating awareness issues in virtual environments have been based on how visual elements can be included in the interface to increase the awareness of others present and what activities are unfolding (see (Gallego et al., 2011) for a recent example). However,
in most of the thesis studies the emphasis will not be on visual elements for acquiring awareness, even though these elements will be considered also. Rather, the focus will be on how collaborating partners can make use of haptic and audio feedback for acquiring awareness information. Mostly, the dialogue between participants in combination with observations of behavior has been the main source of acquiring data about awareness. How the modalities currently supported influence the means of acquiring awareness and the level of awareness achieved is of especial importance in the analyses. Since the ability to generate and feed through relevant and accurate awareness information is of utmost importance for a collaborative interface, awareness has also been used to judge the usability of the collaborative interfaces used within the scope of the thesis.

The main focus in most studies in this thesis has been on action awareness due to the fact that the understanding of what is happening to shared objects during collaborative task solving, if you cannot see, relies heavily on the haptic and audio modalities. Spoken dialogue is of course important too in these cases, but as we will see later on in the studies with visually impaired or blindfolded participants, you can hardly rely on someone to constantly verbalize what is going on during collaborative task solving. Apart from action awareness, social awareness has also been considered to some extent.

Common ground also tends to be defined differently in different studies. The definition given by Clark and Brennan (1991) – a state of mutual understanding among conversational participants about the topic at hand – has been used in the thesis. Again, the focus lies mainly on how haptic and audio feedback can be used to achieve a consensus and especially how the means of establishing common ground changes when the combination of modalities supported changes.

Although relevant, the taxonomy used by Roque and Traum (2008) has not been used, since the analysis of the observation material gained from the thesis studies has not been performed according to that level of detail (see chapter 3 for a more thorough discussion of the analysis method used). However, the taxonomy is highly relevant for the understanding of the results and it will therefore be used later in this thesis when discussing the results in more depth.

The concept of deixis forms one of the cornerstones in the studies and has been considered in great depth during the analyses. Both verbal references and gestures have been studied in detail. The use of deixis has also been the most important element considered when drawing conclusions about common ground. Deictic references in the dialogues between the collaborating participants, in the thesis studies, have also been used as a means of judging how haptic feedback influences these dialogues. Later on in this thesis the deixis concept, originally defined in relation to verbal communication, will be expanded to haptic communication (i.e. haptic deixis). Haptic deixis is especially important when considering the means of utilizing haptic feedback for providing guidance.

### 2.2.3 Collaborative multimodal virtual environments

As is clearly the case for single-user multimodal environments, having access to more than one sense also comes with a lot of virtues in collaborative environments. When collaboration is concerned, there is an extra challenge that lies in developing appropriate
support for establishing common ground and acquiring awareness information, especially if visually impaired users are part of the target group. In that case haptic and/or audio feedback may be necessary for reaching common ground and acquiring sufficient awareness information, provided that all information about actions taken, shared resources and task status cannot be conveyed verbally. Although this section will touch on the subject of interfaces for visually impaired users, most of the discussion about such interfaces will be brought up in relation to the thesis studies about collaboration between sighted and persons who cannot see in chapters 4 and 5.

In the remainder of this section the virtues of using multimodal collaborative environments will be discussed. For clarity, the different feedback combinations visual/audio, visual/haptic and visual/haptic/audio will be separated. The distinction between these cases has also been made in the thesis studies. Of course, collaborative interfaces providing only visual feedback are still the most common but these will only be covered here in cases of comparisons. The point here is to show why support for more senses than vision is often beneficial.

2.2.3.1 Visual/audio collaborative environments

Studies on the effect of using audio cues, often in the form of audio icons or earcons (see definitions in section 2.1.1), in collaborative virtual environments have been around for many years often showing what one can gain from using audio cues to signal presence of other users in the environment (Baharin and Mühlberger, 2010; Isaacs et al., 2002) and for conveying awareness information about actions taken (Hancock et al., 2005; Gutwin et al., 2011). In relation to the earlier discussion of the importance of awareness for effective collaboration these results are of utmost importance.

The ARKOLA bottling plant simulation carried out by Gaver et al. (1991) is one of the pioneering studies when it comes to showing that the use of sound icons can influence task performance, division of labor and collaboration in a collaborative interface. They simulated a soft-drink factory including different user-controlled stations in which machines controlled different parts of the manufacturing of the drinks, from adding components to bottling up the final mix. The goal was to produce as many bottles as possible as efficiently as possible. Gaver et al. (1991) developed a mixture of sound icons which they called “an ecology of sounds” which, at all times, reflected the state of each machine, whether there was a waste on a station (breaking glass, spilled liquid, etc.) and the occurrence of simple actions like button presses. All sounds were playing continuously and hence the sound environment was complex, like in a real bottle-plant. This is a good example of a sonification, as defined in section 2.1.1. The system was evaluated with pairs of users working with the system for one hour with and without the sound feedback respectively. One of the major challenges was that only half of the plant fitted on the computer screens – division of labor was inevitable.

First of all, the results showed that a series of complex processes represented by different sound icons could be successfully interpreted as being one single complex process – namely the one of monitoring the overall process. The sounds gave an overall picture and deviations from this were easily spotted. Second, the results showed the ability of sound icons to convey semantic information about what is happening. Third, and most important for this thesis, the collaboration was eased in a number of respects when the sound cues were provided. Since the users were working in separate parts of the
plant, controlling approximately half of it each, they had to constantly describe problems verbally when there were no sounds present. At several occasions this was shown to be cumbersome and often the users did not attend to what happened at “the other side”. When sounds were present the users collaborated more closely. When they heard a problem at the other side they pointed it out and often solved it together. In this way the process of filling as many bottles as possible was, of course, eased. This is a clear example of auditory icons influencing collaboration by conveying awareness about actions and events. In the same way the sound icons function as shared reference points (Gaver et al., 1991) for the collaborating partners to focus on and hence the sound icons enabled a continuous establishment of common ground necessary for effective task completion. Overall, much due to the ability of the soundscape to convey awareness information and forming a basis for common ground, the task of filling up bottles was performed more effectively and securely when sounds were present.

After Gaver et al. (1991) performed their pioneering study a number of later studies have also shown how sound cues influence collaboration. In a recent study Gutwin et al. (2011) showed how audio cues could effectively convey awareness information about actions taken in a collaborative interface. They evaluated a shared workspace for drawing and especially how sound cues could make the participants aware of off-screen activity. They compared a visual only condition (with real-time radar views of remote activity) with an audio and audio/visual condition respectively. Even though a remote user was simulated it was clearly shown that the addition of sound cues helped especially in cases where the visual radar view was cluttered with drawn shapes, and when it came to detecting what the other person was doing and where. Gutwin et al. (2011) also conclude, as did Gaver et al. (1991), that it is highly important that the sound cue conveys information about the type of action performed. Thus, the sound cues should not only direct the user’s attention to the radar view but also help him/her to focus on a specific aspect on the, often cluttered, radar view. In this way the sounds increase awareness of the overall situation and as a result increase the task performance. These findings can be generalized to a wide range of collaborative environments in which using only vision could be cumbersome or even impossible. Hancock et al. (2005) also performed experiments on the same theme and most importantly stressed the importance of looking at the relation between group awareness and self-awareness. For example, sound icons that indicate an error made by one user could affect the whole group’s perception of the error rate.

In a more recent study Baharin and Mühlberger (2010) also studied the use of sound icons in a virtual environment, but focused more on attention. They showed that using sound icons for environmental sounds, representing another user, does not necessarily mean that the other user’s presence is being noted – you need to listen to the sound, not just hear it (Baharin and Mühlberger 2010). They ran several experiments and in one of them two users were playing a game, not seeing the same scene. The other user’s actions could be heard through sound icons played from a virtual radio. The game was played one time when the participants did not have to attend to the other person’s actions in the environment and one time when each user should write down and categorize each occurrence of an action made by the other user. It was clearly shown that the feeling of co-presence, measured by questionnaire items, increased when they had to actively listen to the sounds and reflect on them. These results are important since they show that
sounds in an interface might be filtered away, especially if they are present at all times. This is important to have in mind when developing and evaluating the influences that sound cues have on collaboration.

To summarize, sound icons or earcons can give added value in traditionally visual interfaces, especially when it comes to the degree of awareness of the activities being performed in the shared interface. Depending, of course, on the design of the specific cues, one can gain considerably from using sound cues. In the thesis studies the influence the sound icons and earcons have on collaboration will be further investigated in the context of joint task solving between sighted and non-sighted persons (Moll et al., 2010) and task solving in an abstract gaming environment (Moll et al., 2013). Since no studies have been found investigating sound cues’ influences on communication and task solving strategies in these and related contexts the thesis studies will bring research forward in these respects.

2.2.3.2 Visual/haptic collaborative environments
Research on how haptic feedback can make a positive contribution in collaborative environments show positive results in many areas. Among the most pronounced benefits are that task performance is increased (Sallnäs, 2004; Chang et al., 2007), the sense of social presence is increased (Sallnäs et al., 2000) and that the degree of awareness increases (García et al., 2009; Hubbold, 2002) when haptic feedback is added to a visual interface. Adding haptic feedback to visual environments also enables the utilization of specific haptic collaborative functions which in important ways can complement the visual feedback (Oakley et al., 2001; Sallnäs, 2004). Haptic feedback also plays a central role in the actual communication carried out between users in collaborative environments. Since the area of communication in collaborative multimodal environments is so central to the thesis the entire section 2.3 will be devoted to it, and it will therefore not be brought up here.

The term task performance has in itself many dimensions. As is the case for single-user environments haptic feedback has been proven to increase task performance when it comes to task solving time, error rates (or more generally; precision) and the quality of the work performed. Sallnäs (2004) conducted experiments on a few different interfaces in which two users should collaborate. The studies measured both task performance and the sense of social presence. The environments were dynamic in nature and included movable cubes which could be grasped and moved around and in one of the experiments two collaborating users could also hand off objects to each other feeling the joint handling of the objects. The pairs who took part in the experiments tested the applications both in a visual/haptic and a visual only case. The results from the experiments were clear – the haptic feedback increased the task performance, both when it came to time to solve tasks and precision and clearly contributed to the feeling of being present in the computerized environment. These findings have been confirmed in a lot of other later studies on collaborative multimodal interfaces. One example is the study of the roles of sensory modalities conducted by Chang et al (2007). They used an air-hockey game to show that the haptic feedback drastically improved task performance in a game that for many years has only been purely visual. They also reached the same conclusions as Sallnäs (2004) and Sallnäs et al. (2000) when it comes to the sense of presence and togetherness.
Hubbold (2002) investigated how a joint precision task benefited from the ability to utilize haptic cues. In the study two users should carry a stretcher, through a chemical plant including different types of obstacles, between two locations. Both users controlled their own side of the stretcher with a haptic device. Even though it was not a comparative study, using a visual-only case as a baseline, important qualitative conclusions could be drawn. It was shown that maneuvering of the stretcher was fairly easy and accurate when the carriers could feel forces from collisions with objects out of view and pushing and pulling forces from the other user. When moving a joint object, as was the case in Hubbold’s study, one needs to constantly account and compensate for the other person’s forces. It is easy to conclude that the joint maneuvering of the stretcher would be much harder if these forces were not fed through the system (Hubbold, 2002). In summary, the haptic feedback made collaboration easier and provided an important sense of awareness of the other’s actions. When moving objects jointly this sense of awareness is of utmost importance.

Oakley et al. (2001) designed and evaluated a distributed text editor in which each user had his/her own cursor, namely the PHANTOM’s proxy. A few novel haptic functions were included in the editor, like scrolling by pulling against the workspace’s edges, dragging the other user to a certain place and finding the other user (with the help of a magnetic force directed towards other’s proxy). These last two functions are examples of what was earlier defined as haptic guiding functions. The task performed during the evaluation was construction of a large UML diagram and the participants were divided into pairs. Half of the pairs used the haptic feedback and the other half used only the visual representation of the interface. The guiding functions proved especially useful in that they enabled the collaborators to use a larger work space while still keeping track of each other’s work. The participants in the visual-only groups packed everything together so they could see what the other was doing. The haptic functions made it possible to work on different parts of the interface – the participants could always use the haptics to find out what the others were working on or show the others what they were working on themselves. However, one of the haptic functions that made it possible to grab the other person was not used very much. This was probably due to social protocols – you do not just grab someone just because you have something to show. By their study Oakley et al (2001) showed that haptic feedback can be used to ease communication and collaboration in a shared interface and the guiding functions, in particular, have been subject to further investigations and development. The study also indicated that the introduction of haptic feedback functions influenced the interpersonal behavior between the collaborators in that they acted according to normal social conventions. The results also clearly show that haptic feedback added to visual interfaces can provide a whole new means of interaction in virtual environments. The haptic functions will be elaborated on further in section 2.3, where communication through mediated touch will be discussed. One of the main contributions from this thesis lies there, since research on collaborative interfaces has not yet covered the influences that haptic feedback has on the dialogue and task solving strategies in the areas studied.

2.2.3.3 Visual/haptic/audio collaborative environments
In cases where three modalities are supported the same kind of results are obtained as in the visual/audio and visual/haptic cases, but they are often more pronounced. Although
not many studies had been performed comparing modality combinations of three senses in collaborative virtual environments prior to the thesis studies, those performed have shown promising results.

García et al. (2009) compared a purely visual condition to a visual/haptic and visual/audio condition respectively. Specifically, they investigated if task accuracy and awareness could be improved by providing haptic and audio feedback respectively during collaborative handling of a beam. The task was for two users to pick up a beam by pulling on it from each side and lift. They should then move the beam together to a new location. The most interesting issue here is that the task is closely coupled in that both users need to collaborate and keep constant track of the overall state of the beam in order to carry out the task.

Three conditions were used and compared against each other. The first condition was a visual only one in which visual cues were provided to show if the beam was about to fall. The second condition was a visual/audio one in which a warning signal was heard when the beam was close to falling. The third condition was a visual/haptic one in which a vibration was felt if the beam was close to falling down. The vibration became more intense the closer the beam was to falling. One of the measures used to judge the results of using the different feedback conditions was the number of times the beam fell to the ground.

The results showed that task performance was significantly improved for the visual/haptic and visual/audio conditions compared to the visual only one, although no significant differences could be found between the visual/audio and visual/haptic conditions. Even though a case with three modalities was not tested, the results clearly show that both haptic and audio feedback could improve task performance and the accuracy with which the task was carried out.

García et al’s (2009) explanations of the results are maybe even more important. In the visual only case all awareness information about the state of the users and the beam were provided by visual cues in the form of arrows for movement directions and a “!”-sign showing that the beam was close to falling to the ground. All in all, quite a lot of information needed to be attended to simultaneously for the users to be able to solve the task in collaboration. In this way, it was hard to keep track of all the awareness information provided and thus the beam was dropped often even though the means of avoiding it were there. When feedback to another sense (hearing or touch) replaced the function of the “!”-sign the results were improved since the cognitive load on the visual sense was decreased and thus it became easier to keep track of all the sources of awareness. By distributing the awareness information to a sense that was “free” in the sense that the task was, in its original state, a purely visual one, the beam was not dropped as often as when only visual cues were provided. The study illustrates, in a concise way, many of the virtues of using audio and/or haptic feedback – cognitive load is decreased, new awareness channels are introduced and task performance is increased in a number of respects. These findings seem to become even more pronounced when the collaboration is closely coupled.
2.3 Communicating in multimodal virtual environments

As has been shown in section 2.2, it is possible to achieve successful collaboration in virtual environments and often haptic and/or auditory feedback proves to be beneficial for joint task completion especially in closely coupled collaboration. With the increasing interest in collaborative virtual environments, dedicated collaborative functions started to emerge based on haptic feedback. Many of these are in some way communicative, like e.g. haptic guiding. These kinds of haptic functions are of special importance to the thesis studies.

Traditionally, communication has often been defined from an information system’s perspective, implying that a certain message is transmitted from a sender to a receiver through a certain medium (Cragan and Shields, 1998). This view on communication is restricted to a one-way transmission of messages. In contrast, one of the characteristics of communicating haptically, both in real life and in virtual environments, is that the communication is reciprocal. Therefore, communication will here be seen as a meaning-making process, allowing for a reciprocal exchange between two or more participants. In fact, an important part of the communicative actions studied in this thesis consists of spoken dialogues occurring in the process of solving a collaborative task. The medium, multimodal virtual environments, is still of importance, but the definition used in the thesis focuses on communication as a collaborative process.

This section will cover communicative functions of haptic feedback in more depth, since this is one of the most important topics for the thesis. Since audio has mainly been used to provide awareness information and not so much for conveying messages within virtual environments, this section will focus entirely on haptic feedback and its potential for communication. Some new communicative functions have been investigated within the thesis work and several have been evaluated within its scope. Thus, some of the main contributions from the thesis can be found here. After a brief introduction to human communication, with emphasis on the sense of touch, this section will discuss the state of the art when it comes to how haptic functions in virtual environments can be used for communicative purposes. Since the thesis studies depend a lot on dialogues between participants the influence of haptic feedback on the dialogue between people will also be discussed in relation to earlier research in the area.

2.3.1 Touch in human communication

To enable verbal communication a wide variety of grounding techniques can be used, as described in section 2.2.1. Since Roque and Traum’s (2008) work on dialogues referenced in that section covers the most important aspects for this thesis when it comes to dialogue analysis, this section will discuss other topics relevant when studying human communication, with a special focus on the role of touch.

Research on the establishment of common ground in different contexts and on how we usually package different types of information has been conducted for several decades when it comes to verbal communication. However, the same research on the touch modality has only recently been the focus of communication research. Our sense of touch has always played a central role in human communication, since the more physical
aspects of communication are important for human development. Through touch communication a newly born child bonds with his/her parents and through the sense of touch a feeling of trust, devotion and love can be built up and maintained (Gallace and Spence, 2010). As for communication through speech, touch can have many different communicative functions and symbolic meanings (Haans and IJsselstein, 2006) and to be able to reach common ground when messages are conveyed through touch these symbolic meanings need to be shared or agreed upon. Numerous conventions exist for verbal communication, but shared symbolic meanings of touch have not been agreed upon to the same extent – at least not theoretically.

In their meta-study on mediated social touch Haans and IJsselstein (2006) discuss six meanings that touch can serve in communication. First of all, touch could be unintentional (accidental touch) e.g. when two people bump into each other accidently. Often, this type of touch encounter is not seen as conveying symbolic meaning. Second, touch could be task-related in that the touch, or applied force, in itself is used to perform a specific action, like pulling someone in a certain direction. Third, touch could convey positive affection. When it comes to symbolic meanings of touch, this is probably the category that has received most attention especially with regard to experiments on mediated social touch. This category concerns, e.g., touch for giving support, showing appreciation and showing love. Forth, there are control touches meant to change the behavior of someone, by e.g. expressing dominance or seeking attention from someone. Fifth, there are ritualistic touches for regulating interpersonal relations. Handshaking is brought up as an example. Last, there are playful touches which receive their symbolic meanings from the current situation and hence the touch itself does not convey a pre-specified meaning.

While Haans and IJsselstein (2006) focus more on symbolic meanings, other researchers have argued that touch is more immediate and thus should not go through the extra step of encoding and interpretation (Wang and Quek, 2010). The immediacy concerns the notion that we do not generally try to convey a particular message when touching someone – the reception of the touch is immediate. Even though there are different views on the role of touch in communication, Haans and IJsselsteijn’s (2006) view that touch conveys symbolic meanings needed to be interpreted, will be used in this thesis.

Just until a few decades ago communicative touch could only be utilized when two (or more) persons were in close proximity to one another. In this case the intended message, which could be of one of the types listed above, is often easily conveyed provided that the symbolic meaning is known to both. However, nowadays technology makes it possible to communicate by touch at a distance through mediated technologies such as computer haptics.

2.3.2 Communicating by mediated touch

Even though there are always issues with e.g. culture and background knowledge, establishing common ground through verbal and haptic communication and gestures is almost always easier in face-to-face communication than in mediated communication (Clark and Brennan, 1991; Olson and Olson, 2000). Communication through mediated
means, whether it concerns phone calls, email or video-conferencing, always comes with a cost (Clark and Brennan, 1991) and this has to be taken into account when deciding about communication medium. Again, mediated communication through text and voice has been studied to a great extent and several solutions exist minimizing the cost associated with the mediation. However, the costs and means of communication by mediated touch are still quite under-investigated.

During the last three decades research on social touch mediated by haptic technology has been developing, even though the focus of the research has changed considerably during the years. This section will provide a walkthrough of important landmarks in this research, starting with simple 2D 1 DOF interactions and ending with coverage of advanced haptic interaction in 3D virtual environments. Please note that presence is not covered here, despite the fact that it is closely linked to communication. Presence is an entire research field in itself and will be excluded since it has not been considered in the thesis studies.

2.3.2.1 The beginning – interaction based on specialized and limited hardware
One of the first attempts to design an application for mediating social touch was the electric arm wrestling system, with which two users could arm-wrestle over a distance (Haans and IJsselsteijn, 2006). The system is purely mechanic and as for many other early systems and prototypes, like the HandJive (Fogg et al., 1998) for entertainment and simple communication, the Shaker (Strong and Gaver, 1996) and inTouch (Brave and Dahley, 1997) for simple communication between two specially built devices, it is limited in scope and application. Even though these early prototypes started off the research on haptic mediated communication and hence form an important cornerstone of that research, they were developed for highly specialized tasks and the conclusions that could be drawn regarding human perception were limited. Plus, most of the early mechanical systems were not in 3D and have not been subject to formal user evaluations and experiments (Haans and IJsselsteijn, 2006).

These early systems were also limited by the degrees of freedom (DOF) and the number of actions that could be performed. inTouch, for example, was based on two devices containing three rotating cylindrical rollers (Brave and Dahley, 1997). When a user rotates a roller the corresponding roller on the other user’s site moves accordingly. Even though communication does not seem to be restricted to a certain context, the number of messages that can be sent in this way are highly limited.

2.3.2.2 Haptic interaction for sending simple messages in software
Around year 2000 studies were beginning to investigate how haptic communication could be utilized as a complement to other forms of mediated communication in virtual environments. Hence, applications were developed which did not only support a limited amount of haptic messages sent between specialized hardware. One of the earlier attempts of incorporating haptic interaction in these kinds of environments can be seen in instant messaging, which will serve as an example here.

Communication through instant messaging has been around for many years and has traditionally depended on input from the keyboard (mainly plain text). One of the early attempts of using complementary haptic functions in communication was the development of “Contact IM” (Oakley et al., 2002). Contact IM is a, partly asynchronous,
instant messaging program including haptic communicative functions and thus concerns the sending and receiving of haptic instant messages through a network service. Only one scenario was used in which a user could select someone from the contact list, pick up a ball and throw it to the chosen contact. The message was asynchronous in that the ball bounced back and forth on the receiver’s side until it was caught. Although not many messages could be sent in this way the force used to throw the ball, and hence the force felt when receiving it, could be varied.

Rovers and van Essen (2004) continued on the same path and evaluated the HIM (Haptic Instant Messaging) in which haptic icons (or hapticons) were used to send haptic messages to contacts. As stated earlier a hapticon can be defined as “brief programmed forces applied to a user through a haptic interface, with the role of communicating a simple idea in a manner similar to visual or auditory icons” (Enriquez and MacLean, 2003). The HIM is built on the idea that a kind of haptic language complementing e.g. smileys will emerge over time. Some haptic messages, mostly representing emotions, were built-in and it was also possible to design one’s own haptic messages. The messages could be sent either by writing a command in the text chat (and hence the haptic message will serve as a complement to the written text) or by using specially designed hardware. By connecting vibration patterns to specific messages (hapticons) a new haptic language can be formed complementing the textual one. By introducing hapticons in this way one may overcome some of the challenges of missing subtle non-verbal cues in instant messaging (Rovers and van Essen, 2004).

2.3.2.3 2D haptic interaction for direct communication in virtual environments

More advanced haptic functions for collaboration between two users in shared interfaces, like haptic guiding, were also taking form around year 2000. Haptic guiding functions come in many forms and enable physical interaction between two hardware devices through their proxies. This interaction is often mediated through a magnetic force between the proxies. The word “guiding” refers to the fact that these functions make it possible to grasp another user’s proxy and physically guide it to another location. The interesting issue here is that applications involving haptic functions, which can be used as a complement to other forms of interaction in virtual environments, can make it easier to investigate why and how we chose to utilize haptic feedback for communicative purposes. Additionally, these new functions that enable direct contact between proxies enable a lot more complex communication than the earlier attempts with specialized hardware and/or vibration patterns.

One of the first examples of a study including 2D complementary haptic functions in which simultaneous interaction between two hardware devices (through the proxies) were used for communication purposes (in this case guiding) was conducted by Oakley et al. (2001) whose study on shared text editing was introduced earlier. Their study will here serve as the example of 2D haptic communication in virtual environments. They developed five different types of haptic communicative functions based on interaction between the users’ proxies. The idea with these functions was to solve problems with awareness (knowing what the other person is working on in a large document or workspace) and remote gesturing (Oakley et al., 2000).

First of all, the users could feel each other in the workspace by a small repelling force pushing them apart as soon as the proxies came into contact. When using this function it
is not only possible to (almost) physically interact but you can also e.g. convey a message that the other user should move away from the current location. Second, a function for grabbing the other user was implemented, enabling one user to physically move the other to a new place. Thus, by using this function one can e.g. show another user what he/she is doing at a particular place without having to give verbal directions on how to get there. The implications of this will be discussed further below. These two haptic functions first and foremost support remote haptic gesturing in that they are explicitly used to direct attention to specific places or directions.

The third function implemented gave awareness information about the other person’s whereabouts in the workspace without any need for visualizing this information. A vibration applied when the other user gets in close proximity to one’s own proxy or a small resistance felt when one gets close to the other user gives a sense of proximity which is felt by both users. This is a kind of awareness update function which is not driven by the users, but it still serves an important social purpose and clearly illustrates the potential of haptic functions in shared interfaces. Note that this function, which is first and foremost a social awareness function, could be used for communicative purposes by e.g. “pushing away” the other user.

The fourth function implemented by Oakley et al. (2000) was also a kind of awareness supporting function which enabled one user to be physically dragged to the other user’s current location by a constant force. In this sense the function did not provide a continuous awareness of the other’s whereabouts but rather served as an awareness update. This function can be used to locate the other user and find out what he/she is currently working on. The fifth and last function implemented was also a locating function, but in this case the other user was dragged to one’s own proxy.

As stated earlier especially this last function and the grab function were not used very much in evaluations because of the intrusion in the other’s work and privacy (Oakley et al., 2001) – you do not just grab someone and move them to a place of your choice. That social protocols are being adhered to in this way highlights the consequence of using haptic feedback for communication in virtual environments.

2.3.2.4 3D haptic interaction for communication in virtual environments

Although the application developed by Oakley et al. (2001) is interesting in that it seems to be the first study which enables close to physical interaction between proxies in virtual environments, the communication was mainly in 2D and the haptic functions implemented displayed constant forces. The studies by Sallnäs et al. (2000) and Sallnäs and Zhai (2003) referenced earlier also included special haptic functions for communication. These are in 3D and thus enable even more versatile means of haptic communication. These functions are also continuous, in contrast to the constant force used in (Oakley et al. 2001), in that the proxies directly influence each other through attracting forces as long as the functions are used.

Especially one of the functions including a function for hand-off of objects is particularly interesting to consider here since it enabled direct interaction between two hardware devices, as was the case for the functions implemented by Oakley el al. (2000). The hand-off study (Sallnäs and Zhai, 2003) added a new collaborative function to the repository by enabling joint manipulation of objects. Each object present in the interface could be picked up by any user by pushing a button on the haptic device while touching
the object. If the users picked up the same object, as was necessary when the object should be handed over from one user to the other, both users could feel each other’s pushing and pulling forces on that shared object. The communicative function in particular was not evaluated in depth, apart from through a questionnaire on social presence, but again the potential of using haptic communicative functions in virtual environments is highlighted. Similar findings were also generated by Hubbold (2002) in the joint stretcher carrying study referenced earlier – adding a haptic feedback channel of communication in joint object manipulation tasks definitely influences both the collaboration and the task outcome. Apart from showing that haptic feedback has a clear effect when handling shared objects, Sallnäs et al. (2000) also showed that one could use the haptic feedback to lift objects while pushing against them from opposite sides. Even in this case the force felt from the other person gave valuable information necessary for effective object handling.

The guiding functions developed and evaluated by Sallnäs and Zhai (2003) are worth elaborating on further here, since variations of those functions have played major parts in some of the thesis studies (Moll et al., 2010; Moll and Sallnäs, 2013). By introducing haptic feedback in virtual collaborative environments new types of collaborative functions can be developed that connects the two hardware devices, as described earlier. In this thesis these function are called haptic guiding functions. One special aspect of haptic interaction is that it is reciprocal in the sense that you receive feedback on your actions at the same time as you provide feedback to the interface or the other user. Applied to haptic guiding functions this means that you can communicate information to the other user, haptically, at the same time as you receive haptic information from him/her. Enabling interaction between two hardware devices thus make new means of interaction possible by which close to physical interaction can take place as was the case when the handling of shared objects was being felt by both users (Sallnäs and Zhai, 2003; Sallnäs et al., 2000). This section has covered a few different kinds of haptic functions for communication. As will be shown later on in this thesis the most important distinction, when guiding is concerned, is the one between forced and continuous. The symbolic meaning of these two alternatives is quite different. To be able to use continuous guiding functions the proxies necessarily need to be in close contact and in this way the guiding procedure is more immediate than if a constant dragging force is applied. Additionally, the continuous guiding enables reciprocal communication in that both persons can apply forces simultaneously. If the guiding is constant the guiding is “forced” in that the destination and trace of movement is already decided when the guiding starts. The distinction between these two cases will be elaborated on a lot more in later chapters.

In the thesis studies haptic guiding plays a very important role and then especially in the studies concerning interaction with sighted persons and persons who cannot see (either visually impaired or blindfolded). In these cases it will be shown that using haptic feedback is necessary for both inclusion of the visually impaired/blindfolded user and for the users to achieve a common ground (Moll and Sallnäs, 2009). Haptic guiding functions and the results of their reciprocal nature will be elaborated on a lot more, further on in this thesis, in relation to these studies (Moll et al., 2010; Huang et al., 2012; Moll et al., 2013). Earlier studies have not investigated the influence that these functions have on collaboration during joint task solving and in this respect the research presented in this thesis will add to the research on the potential of these types of functions.
2.3.3 Haptic feedback can influence the dialogue

The study referenced earlier performed by Oakley et al. (2001) also shows how haptic feedback could influence the dialogue. Using the haptic guiding functions it was easy to find the other person or show the other person interesting parts of the interface. In the visual-only groups, however, the dialogue often focused on verbal guiding either by direction words or by referring to parts in the interface. This was of course cumbersome and time consuming, since a person who may not see the other person’s position or feel anything had to be verbally guided to different specific details in the interface.

Another study that shed light on how haptics influences the content of the dialogue is that conducted by Bivall Persson et al. (2007). They developed and evaluated a haptic feedback system for protein docking, intended for use in educational settings. A typical task performed with the system could be to dock a ligand with a large protein structure. Forces between the ligand and the protein should give the user a cue of when and where a docking should be made. The evaluation was performed in a lab with students studying biodynamics. Results showed that docking tasks were made faster with haptic feedback, even though the quality of the result did not seem to be affected by the addition of the haptic force model. However, interviews revealed that the haptic feedback helped the students see “the whole picture” with different forces and dynamics. Another important result was that students talked more about forces and dynamics than usual after using the system.

The same haptic feedback system for ligand docking was evaluated by Tibell et al. (2009), with two groups of students; one using the haptic feedback system and the other one using only the visual presentation. This time the results showed more clearly a positive effect on learning but more importantly the results again showed that the haptic feedback influenced the reasoning about the dynamics of ligand docking considerably. Grasping the concepts of forces between molecules and ligands is known to be hard and this makes the result of the evaluation even more promising, since it clearly shows that the students who used the haptic feedback could reason about these concepts. Similar results were gained by Chellali (2011) who developed and tested a system for learning needle insertion. In the collaborative system, one user guided the other by a function similar to the ones evaluated by Sallnäs and Zhai (2003). When comparing a visual and visual/haptic version it was clear that the haptic feedback influenced the dialogues. Those who used the haptic version of the training application made more references to haptic sensations and fewer references to places and directions.

It is clear that haptic feedback can influence the dialogue and the studies referenced above show, first and foremost, that the focus of the discussions changes when haptic feedback is provided. Oakley et al. (2001) also touches on the subject of guiding, which will be further investigated in the thesis studies discussed in chapters 4-6. As stated before the thesis studies will deepen the insights into how haptic and audio feedback influences the communication between two persons engaging in task solving and also take a step further and use these insights to explain quantitative task performance results.
3 Methodological concerns

In this section the methods and metrics used for analyzing the different studies performed within the scope of the thesis project will be elaborated on. The purpose is to provide a theoretical framework for the methods used. The focus will be on the dialogue analysis as a means of evaluating collaboration.

3.1 Analyzing dialogues

Dialogues between participants have played a major part in the analysis of the results from all studies performed within the scope of this thesis. In this section, some of the most commonly used methods for analyzing dialogue data – content analysis and conversation analysis – will be gone through and related to the methods used in the thesis work. First, these two methods will be theoretically introduced after which the method used in the thesis studies will be explained and related to the theory.

3.1.1 Dialogue analysis approaches

Content analysis, based on an emergent coding scheme, is the type of analysis that is closest to the method used in the thesis studies. Using content analysis is often an efficient and replicable way of arranging large amounts of transcribed material into well-defined categories (Stemler, 2001). The method is general in that it can be used on any kind of written content, thus it is not restricted to analysis of transcribed material. When content analysis is applied a representative part of the material is typically gone through by two independent researchers each producing a list of suggested categories (Stemler, 2001). After consolidating the lists the researchers then use the identified categories on a small part of the material. The results of the coding are then compared and if they pass a reliability check, the coding is used on the entire material. Lantz (1993) describes a similar method used when analyzing interview results. The method focuses on coding data according to identified dimensions (or categories as Stemler (2001) calls them) reflecting different aspects of the content analyzed. After arranging the coded data into the different dimensions the researcher then looks for patterns both within and between the dimensions. Lantz (1993) also acknowledges that the reliability of the analysis can be judged by having an independent researcher redo the analysis, but does not highlight it as a mandatory step of the analysis process.

Another widely applied method used for analyzing dialogues/transcripts is conversation analysis (Greatbatch et al., 1995; Pomerantz and Fehr, 1997). There are several aspects or assumptions that found the basis for conversation analysis and which in part differentiate it from other methods (Greatbatch et al., 1995). First of all the language use per se is not of main interest when performing this kind of analysis. The language is rather seen as a tool that conversational partners use in order to negotiate meaning and produce relevant actions. The context plays a central role in conversation analysis. When applying conversation analysis, however, you do not look at context as something
external but rather derive the context from the interaction and thus you only look at context aspects that are relevant to the conversational partners. Thus, the context in itself “consists of and is reflexively constituted in and through the participants’ actions and activities” (Greatbatch et al., 1995). In this sense it is also clear that the context is never stable, it is constantly renewed as the interaction proceeds (Pomerantz and Fehr, 1997).

Third, the order in which utterances line up in a conversation is of utmost importance and one of the basic analysis elements. When performing conversation analysis of a dialogue excerpt, you do not only look at the dialogue as a whole or each utterance in isolation but also at the relations between consecutive utterances.

Pomerantz and Fehr (1997) describe five main steps which are, more or less, performed every time the conversation analysis approach is used to analyze spoken interaction: 1) Select a sequence, 2) Characterize the actions in the sequence, 3) Consider how the speakers’ packaging of actions, including their selection of reference terms, provides for certain understandings of the actions performed and the matters talked about. Consider the options for the recipient that are set up by that packaging, 4) Consider how timing and taking of turns provide for certain understandings of the actions and the matters talked about, 5) Consider how the ways the actions were accomplished implicate certain identities, roles and/or relationships for the interactants.

### 3.1.2 Dialogue analyses in this thesis

A content analysis approach based on an emergent coding scheme has been used in the thesis studies and the list of categories has also been iterated before being applied to the entire material, including the interview results. In this way the analysis of the transcriptions has followed the approach discussed by Stemler (2001) aside from the fact that only one researcher (me) has derived the categories and applied them to the material. After each iteration of the category list and after the category list had been applied to a part of the material, thorough discussions with research colleagues occurred to ensure a high validity.

In a first step of the analysis process the recordings from all evaluation/experiment sessions were transcribed in their entirety. Both communication between participants and observations of behaviors were annotated. Based on an explorative coding approach categories were then defined and applied to small parts of the material during a first examination and hence no categories for coding the material were decided on before the coding began. Thus, an emergent coding scheme was always used to derive categories during the first phase of the iteration of the transcribed material. Some of the categories used in the coding, like Social awareness and Common ground directly related to theory. Therefore it is worth pointing out that no categories were decided on before the analysis began – these kinds of categories were inspired by theory but they would not have been used unless phenomena directly related to the definitions of these theoretical terms had arose in the material considered. After the initial coding the categories found were formally defined and discussed with research colleagues. After 2-3 iterations the resulting list of categories was then used to code the entire material. Note that e.g. an utterance or a short dialogue can fall into several different categories depending on the content of the utterance/dialogue. As an example the following short dialogue (extracted from a
transcript from the study presented in chapter 6) was assigned the categories Referencing and Gesturing

P1:  *Here!*  
     [P1 indicates the place by moving his avatar up and down]

P2:  *Ok, there!*

Note that the entire dialogue was coded in this case, but not each single utterance or observation. After the entire material had been coded the data units were sorted under their respective categories for each participating group respectively. These were then compared across experiment conditions and groups. In a last iteration of the data material dialogue excerpts (like the one shown just above) were extracted, which served the purpose of illustrating the most central findings. Extracting these dialogue excerpts was an important part of the analyses since the dialogues could be used as a basis for discussion about main findings. According to Korolija (1998) episodes in conversations have; (1) defined end points, (2) some kind of thematic and/or action unity, (3) a participation framework (e.g. a collaborative activity) and a (4) core event structure. The focus in all episodes, presented in dialogue excerpts in this thesis, is the communicative activity of solving a particular sub-task (e.g. moving a cube or reach consensus about a task solving strategy). Note that the description of the analysis method given here only focuses on the purely qualitative analysis. Several of the analysis results from the thesis studies were re-examined to derive quantitative data. That process will be described in section 3.2.

The conversation analysis approach has been used on parts of the transcribed material. When following the conversation analysis approach strictly each single utterance, and pauses between utterances, should be coded separately on the *entire* material. In the last phase of the analysis, when dialogue excerpts had been chosen several of the steps applied in conversation analysis have been used for analyzing those particular dialogues. When for example looking at how different actions depend on each other and how the participants in the experiments e.g. establish a common ground the method applied during the thesis follows the conversation analysis approach. This part is very important if the haptic or audio feedback is the only means of establishing a common ground, i.e. when sight is not available to one of the participants. Utterances from visually impaired or blindfolded participants like “this one is longer” and “I’m in the corner” highlight interesting aspects when it comes to haptic communication and thus these kinds of verbal references have been considered in depth. When it comes to taking pauses and non-verbal events into account the conversation analysis approach has been used in that all of those are considered in order to identify whether the use of haptic or audio feedback might have caused them. For the particular dialogue excerpts chosen the pauses (defined as the time between two consecutive verbal utterances) were exactly timed, after a re-examination of the recordings. The lengths of these pauses were, however, not annotated in the excerpts chosen. The dialogue below, taken from the study presented in chapter 4, illustrates a situation where special attention to pauses proved to be of importance.

**Sighted:**  *All right, you can pick that one you have now*  
     [The visually impaired pupil picks up the cube]
Sighted:  *And then, ..., wait, ..., a little bit more to the right.*  
*Up, take it up a little, ....*  
[The visually impaired pupil moves up towards the roof]  
...

Vis. impaired: *No, it does not work*  
[The sighted guy picks up the same cube to help]  
[They move the cube towards them]  
[They move the cube a tiny bit to the left]  
[They place the cube to the right of another one]  
[They fine-tune the cube’s position]

Sighted:  *That is good!*  
[The visually impaired and the sighted pupil let go of the cube]

If special attention was not paid to pauses in the *verbal* communication it could be easy to miss the fact that there was a very important pause, for about 5 seconds, between the last two utterances shown above. Note that the pause is only indirectly indicated in the example, as the series of notes about what is done between the two last utterances.

Without considering the pauses one of the most important findings from the studies might have been missed completely – haptic feedback can replace the need for giving verbal directions (Moll and Sallnäs, 2009). This is just one example when attention to detail can make a huge difference and in this respect the approach followed for analyzing the dialogue excerpts is similar to conversation analysis. It should be noted, though, that e.g. the pauses were not exactly timed in the entire transcribed material and that the length of particular pauses are only reported in the thesis in a few cases. The example shown above will be more thoroughly treated in section 4.6.2.

A total of 16 dialogue excerpts have been used to illustrate main findings throughout this thesis. Even though these particular excerpts present about 1% of the entire transcribed material from all studies, care has been taken to use examples which are representative for large parts of the corpus material. These examples are defined according to the four points stated by Korolija (1998). In the example just provided, the episode illustrates the act of helping the visually impaired pupil to position a cube in a particular place – the episode starts when one of the sighted pupils initiate the guiding process by referencing the cube and providing a verbal direction and ends when consensus is reached about that the cube has been correctly placed.

### 3.2 Quantitative and/or qualitative data?

As discussed in section 3.1, dialogues have played a major part in the analysis of the results of the studies performed. Dialogue excerpts have been used to illustrate the main findings. While this often proves to be awarding it has a downside – the dialogues are qualitative in their nature and as such they are always subject to interpretation. The researcher plays a role both in the interpretation and in the selection of excerpts and findings to present.

Even though formal procedures have been followed during the qualitative analyses performed, quantitative measures have been used in all studies as well. These measures...
do not only concern data that were subject to statistical analysis but also data collected by quantifying qualitative results. The transcripts of the video material generated from the thesis studies have undergone several analysis iterations. At the outset of the first studies performed (see chapters 4 and 5) only qualitative variables were used for evaluating the collaboration between study participants. At later iterations – after re-examining the material during several publication processes – the conversations and actions carried out were also studied in order to derive quantitative variables. One example is the incidence rates of different kinds of guiding actions observed (visual, haptic or audio) in the study comparing a visual/haptic interface and a visual/haptic/audio interface. Quantifying observational data turned out to be rewarding and this is why quantification of qualitative data was decided upon at the outset of the last collaborative study performed (see chapter 6). The quantitative measures have been used to strengthen the reasoning about the qualitative results gained from the thesis studies.

One last method utilized in all thesis studies is the semi-structured interview, also subject to the content analysis method described in section 3.1, often rounding off the evaluations or experiment sessions. The interview results further helps in the interpretation of observational data since they present the participants’ own views of what went on during collaboration. This is data that could not possibly be gathered only through observations or by analyzing quantitative data and hence they add a new dimension to the results.

To conclude, the analysis of dialogues, the statistically analyzed data, other quantified measures and interviews have all been used for triangulation in the studies presented in chapters 4-6. All of these data gathering methods combined provide a thorough foundation for the conclusions presented in the thesis. With this in mind, we will now move on to the first study performed within the scope of the thesis work concerning collaboration between visually impaired and sighted pupils.
4  Haptic feedback for collaboration between visually impaired and sighted pupils

In the following chapter the first thesis study will be introduced and elaborated on. This study focused on collaboration between visually impaired and sighted pupils in schools. The rationale behind this study, which will be discussed thoroughly in this chapter, was to investigate how a visually impaired person could get included in collaborative work and then especially how haptic feedback could provide the common ground and awareness necessary for the inclusion to be possible. Overall, the study yielded important insights about the haptic feedback’s role in this very special collaborative situation and fruitful ideas and important research questions for the next follow-up study. This study was unique in that no one had investigated haptic feedback for joint object manipulation in a similar setting prior to it.

The applications tested were developed by me and were based on earlier applications evaluated by Sallnäs (2004). Eva-Lotta Sallnäs Pysander and I performed the user evaluations together on the study sites. I also performed the qualitative analysis of the data material, which has been iterated during the recent years as new insights were gained from later studies. Eva-Lotta and I collaborated in writing the resulting article (Moll and Sallnäs, 2013) and the conference papers (Sallnäs et al., 2006; Sallnäs et al., 2007).

An example view of the test setting in one of the schools.
4.1 Group work in Swedish elementary schools

It is common knowledge that group work is getting increased attention in Swedish schools. It is even stated in the school curricula that all pupils, including visually impaired ones, should take part in group work activities. Group work not only promotes learning, but also forms social ties between pupils. Therefore it is of utmost importance to support inclusion of all pupils in such activities.

In the thesis study presented in this chapter the school context was chosen since material from an earlier field study (Sallnäs et al., 2005), discussed in section 4.1.2 below) showed that there was a severe lack of inclusion of visually impaired pupils in group work in elementary schools studied. The hypothesis was that getting access to the same shared haptic interface would aid the visually impaired pupil’s inclusion in group work activities. The work was carried out within the scope of the EU-project MICOLE (Multimodal Collaboration Environment for Inclusion of Visually Impaired Children) which aimed at developing and evaluating tools for supporting inclusion of visually impaired pupils in group work activities in schools.

In this section theories related to learning and the results generated from the field study (Sallnäs et al., 2005) will be discussed. Since learning is not one of the main topics for this thesis the theoretical discussion about learning perspectives will be kept brief. More information can be found in the references sited and the article (Moll and Sallnäs, 2013), which also includes more contextual information. The emphasis in this thesis will be on how haptic feedback influenced the collaboration between the pupils.

4.1.1 Some theoretical points about learning

The socio-cultural perspective is a pedagogical perspective receiving a lot of attention in today’s education. This perspective states that pupils actively construct their knowledge of the world by communication with peers and manipulation of common artifacts (Vygotsky, 1978). According to Säljö (2000), development in this perspective equals “a socialization into a world of actions, conceptions and social interplay that are all cultural”. It is not hard to see that the school and especially the group work method plays an important role in teaching and in creating social ties. It has also been established that the pupil’s reasoning and understanding is increased and that more complicated problems can be solved when working with peers than when working alone (Decortis et al., 2003).

Storch (2002) reports on an extensive observation analysis of group work between pairs of adult students during an entire semester. The issue of focus during the study was the nature of interaction between the collaborating peers. The tasks, all related to writing academic texts, were presented during a language course as any other learning material. According to the results of this study two persons collaborating on a joint task can take on different roles relative to one another. The four patterns identified were; collaborative, expert/novice, dominant/dominant and dominant/passive. From the observation analysis it was clear that “collaborative” was the pre-dominating pattern. One of the important insights was that more transfer of knowledge was shown to occur in pairs working collaboratively. Storch (2002) shows that it is important to consider the group work situation and the roles taken when evaluating the efficiency of work. Thus, supporting a
collaborative atmosphere is just as important, and maybe more so, than providing the opportunity to do group work.

When considering the situation for visually impaired pupils integrated into ordinary classes this discussion has important implications. As will be discussed more thoroughly in the next section, it has been shown that visually impaired pupils often take the passive role in quite a few group work situations (Sallnäs et al., 2005). In light of the results gained by Storch (2002) this implies that the visually impaired pupil does not gain as much from the group work as sighted pupils do. This is one of the reasons why it is important to create tools to support inclusion of visually impaired pupils in group work situations in order for them to be able to take part in the collaborative work on equal terms. Computer technologies may support the creation of a joint problem space and a shared frame of reference, needed by pupils engaging in group work. Related to this, Gibson (1979) stresses the fact that the affordances of different interaction media and their effects on collaboration need to be considered when aiming to support a fruitful and effective group work. This has to be taken into account when developing group work interaction tools and this is especially important when developing technology used by both visually impaired and sighted pupils.

4.1.2 Differences in work material

Within the MICOLE project Sallnäs et al. (2005) performed a field study in schools in several countries to investigate how group work involving visually impaired pupils is carried out. Observations of group work were performed and interviews were conducted with both pupils and their teachers. Several problems were identified when it came to inclusion of the visually impaired pupil and the possibilities for visually impaired and sighted pupils to reach common ground about available resources and the task at hand.

First of all, interviews with teachers and personal assistants\(^4\) revealed that it was generally harder for the visually impaired pupil to access material. Material could be accessed by the visually impaired pupil through screen readers, braille, etc., but reading by means of such equipment took a lot of time. This has consequences for division of labour and often limits the roles that visually impaired pupils can take in group work. Partly as a consequence of this, the visually impaired pupils interviewed did not generally see group work as rewarding as the sighted pupils did.

Second, the difference between available work materials in itself often caused difficulties during group work. One of the most basic problems included writing texts and even short notes. Despite the obvious fact that it is cumbersome to write long text segments on a Braille computer the whole text cannot be viewed by the sighted pupils when notes should be compared or complete reports composed. Thus, the available materials for writing and reading reports or notes are generally not adapted to collaboration. This problem was noted by both visually impaired and sighted pupils. The fact that reporting tasks could not be performed on equal grounds and that the material produced could not be shared in a straightforward way of course constrained the possibilities to collaborate and communicate since it was hard to reach common ground.

\(^4\) A person that is employed by the school to help the visually impaired pupil in the daily work. The personal assistant e.g. escorts the visually impaired pupil around and prepares work materials.
Third, the difference in work materials also caused problems with inclusion of the visually impaired pupil in other parts of the work. It was often seen that visually impaired and sighted pupils worked completely in parallel in separate work processes much due to the difficulty of achieving common ground based on the available work material. Apart from the problem with presentation on Braille displays most of the material the visually impaired pupil had access to was by some means specially prepared by the personal assistant. It was not certain that the visually impaired pupil would get access to material at the same time as the sighted pupils and when he/she did it was hard to share the information. Since blind persons cannot access mathematical formulas and illustrations by means of vision, special techniques must be used such as special Braille notations, raised line drawings and clay ropes. Thus, the means of accessing and sometimes also developing content differs completely between visually impaired and sighted pupils.

The issues identified during the field study (Sallnäs et al., 2005) were problematic in that they caused exclusion of the visually impaired pupil from important information, and thus one of the major benefits of doing group work was constrained. The aims of the MICOLE project was to develop tools to aid the inclusion of visually impaired pupils in the group work situation and the applications developed in the study presented in this chapter are two examples of the resulting tools. Before the applications are described and the results are discussed other examples of software support for visually impaired persons will be discussed.

4.2 Software support for visually impaired

Research has been conducted during several decades regarding the use of technology to support visually impaired persons. The most important challenge that lies in developing computer interfaces for that target group is that all information has to be transmitted through other senses than vision. Deep insights into how visually impaired persons explore the world are required and also knowledge about plasticity and cross-modal relations. Relating to the discussion of common ground and grounding in section 2.2.1.2, it is even more challenging to develop tools for group work between sighted and visually impaired persons since the representations provided to the sighted and visually impaired user are usually different. This section will only focus on software which relate to the thesis study in that they are used to support learning and/or collaboration between visually impaired and sighted people. Note that collaboration between sighted and visually impaired users was chosen as a research area in this thesis due to the potential of gaining insights of how haptic feedback functions can be used for communication generally. Thorough studies of support for visually impaired users or their perception of the world lies outside the scope of the thesis.

4.2.1 Designing haptic interfaces for visually impaired users

Since development of haptic interfaces for visually impaired persons is particularly challenging, since these users cannot make use of complementary visual information, design for this user group will be specially treated here. Other issues related to design of
multimodal interfaces will be treated when relevant when the thesis studies are discussed. The design criteria brought up here are based on work by Sjöström (2002).

First of all it is important to create a free-standing virtual haptic object design, which does not rely on visual input. It is especially important to render all sides of the objects (not just the ones you can see) and edges should also be smooth so objects are easy to follow. As a side note to this design criteria it is central to point out the importance of having a graphical interface which is as equivalent as possible to the haptic counterpart when visually impaired and sighted persons should collaborate, as in the study described in this chapter. This correspondence between haptic and graphical feedback is important for the establishment of common ground, as discussed earlier. Since the designer is often sighted extra care must be taken to meet this design criterion since important details can easily be missed due to visual dominance (see the discussion about cross-modal relations in section 2.1.2).

Second, navigation and overview should be supported. There should be a lot of reference points in the interface, which the visually impaired person could refer back to if he/she gets lost or want to remember positions of specific key objects or interface elements. Getting an overview solely based on the sense of touch is much harder and time consuming than getting an overview by looking, and this fact should be reflected in the interface design. For example, the complexity of the interface should be kept to a minimum. Since the haptic exploration is constrained when using a probe (Lederman and Klatzky, 1999) extra attention must be devoted to these design criteria when developing interface based on one point interaction through e.g. a Phantom.

Third, it is important to use all accessible modalities, so that the touch modality will not be overloaded. One possibility, often overseen, is to use sound cues of different sorts to give more concrete information about details or data values. Several of the examples provided in the next section show an added benefit of using sound cues in multimodal interfaces for visually impaired persons.

Apart from the above guidelines the usual guidelines and principles related to usability of course also apply when developing interfaces for visually impaired people.

4.2.2 Multimodal environments for visually impaired users

In the MICOLE project introduced earlier several other applications, apart from the ones developed in the thesis study presented in this chapter, were developed aiming at supporting inclusion of visually impaired pupils in group work settings (Winberg, 2006; Magnusson et al., 2007; Rassmus-Gröhn et al., 2007; Archambault et al., 2007; Plimmer et al., 2008). Winberg (2006) developed and evaluated a cross-modal system for direct manipulation to find out if a sighted person using a visual display and a visually impaired person using an auditory display could cooperate on moving elements using the familiar drag-and-drop metaphor. The interface consists of four quadrants in which objects are placed (see the circles in figure 6 below).
All objects in the interface have separate sound beeps associated with them. When the pointer is placed in a quadrant by the visually impaired person all objects in that quadrant are played in sequence. The volume of a beep tells the user how far the proxy is from the respective objects and a kind of guide tone also reveals if the proxy is above, below or on the same horizontal level as the respective objects. The users interact with the interface using a pen stylus on a graphic tablet and by using a button on the pen the users can pick up and move the objects. This interface was evaluated in a collaborative study in which pairs including one sighted and one visually impaired person participated. To enforce collaboration the sighted user could only see the content in two of the quadrants and the visually impaired one could only hear the content in the remaining two quadrants. Thus, the users needed to collaborate based on different representations of the same information, to be able to solve the task. During the evaluation the participants were asked to solve tasks which aimed at sorting a number of objects by moving them between quadrants. The results show that the interface was accessible in that the visually impaired user could get access to and use all available information. The application also promoted collaboration and inclusion in that both could do an equal amount of work (Winberg, 2006). However, one important finding relevant for task design was that it took longer time for the visually impaired person to get an overview of the contents in the quadrant than for the sighted person who could just glance quickly. This gave the sighted person an advantage since he/she could start thinking about task solving procedures faster. Thus, even though the interface promoted collaboration and inclusion the tasks could not be solved on equal grounds. The differences between our senses’ capabilities need to be taken into account when designing tasks. The most important finding for this thesis, however, is that the two users could collaborate on solving the tasks despite the fact that they had access to different representations of the interface. Winberg and Bowers (2004) also presented a similar finding in their study on the Towers of Hanoi puzzle introduced in the theoretical background, showing that pairs of visually impaired and sighted users could solve the puzzle based on an auditory and visual representation respectively.

Magnusson et al. (2007) and Rassmus-Gröhn et al. (2007) developed and evaluated a multimodal drawing application based on visual, haptic and audio feedback. The application was initially designed to make it possible for visually impaired users to create 2D relief drawings by means of PHANTOM devices (Rassmus-Gröhn et al., 2007). The
motivation was to present a way to create and explore simple graphics without the need for equipment specially prepared for the visually impaired pupil. While the haptic feedback is mainly used for drawing and exploration, the audio feedback is used to playback verbal description tags on images currently touched. A mouse is also added to the system, which can also be used to draw and explore images but then without haptic feedback. Additional keyboard commands could also be used to e.g. tag and delete lines. The most important feature for the mouse is that the mouse user can drag the PHANTOM user to the mouse pointer’s position by clicking one of the mouse buttons. The PHANTOM user can do the same thing by pushing the button on the device handle. In this way the mouse user in particular can guide the PHANTOM user to a certain location. This guiding function is very similar to the “come-to-me function” developed by Oakley et al. (2001) in their shared editor system. This is important since this shows that haptic guiding functions can serve purposes in both interfaces for sighted users and interfaces for visually impaired ones. Most importantly for the MICOLE project the drawing application was evaluated, by Magnusson et al. (2007), in schools with pupils using it to perform ordinary school work. The task was to discuss composite drawn figures using mathematical terms like “sphere” and “angle”. The pupils worked in groups of two and the task was for one pupil to first draw a composite shape without showing it to the other pupil. Then, the pupil who drew the shape should describe it to the other, who should try to mimic it. In the last phase the two pupils should compare the two drawings, again by using a mathematical language. Figure 7 below shows the drawing interface and a few examples of drawn figures/replicas.

Since the evaluation was performed in a class including a visually impaired pupil, this pupil and one of the sighted pupils did the task in the drawing application. The results showed that the group using the drawing application could indeed solve the task of replicating each other’s drawings. Most importantly for this thesis it was shown that the guiding function, used when a sighted person wanted to show his/her replica to the visually impaired one, was used successfully. An important distinction was also made between (1) positioning the mouse close to the PHANTOM proxy dragging it towards and around the figure and (2) positioning the mouse by the drawn figure dragging the PHANTOM user in a straight line towards the figure (Magnusson et al., 2007). The second option was seen to be more artificial since it did not convey a real sense of guiding but rather a feeling of a constant force dragging the user in a pre-defined
direction. Plimmer et al’s (2008) study on teaching handwriting to visually impaired pupils resembles the drawing application just described. They developed a system which allowed real-time playback of shapes drawn by a teacher. Characters could be drawn by the teacher on a tablet-PC and at the same time the visually impaired pupil could follow the teacher’s movements by holding on to a PHANTOM device. It was also possible for the pupils to draw shapes on their own. An initial evaluation showed that the system could indeed be used to train visually impaired pupils to draw simple characters.

Archambault et al. (2007) took a slightly different approach in that they tried to fuse the technology used by visually impaired and sighted users respectively to create a shared interface. They developed a model for presenting mathematical expressions to both Braille displays and ordinary computer screens for visual output. Their work is important for inclusion of visually impaired children in school work since mathematics is seen as one of the hardest topics for them and requires material completely different from the ones that visual pupils use (Sallnäs et al., 2005). The two views (Braille and visual) are synchronized in that selections of particular parts of a formula and updates to a formula are both seen and felt. The visually impaired pupil can also “point” when at a particular spot in the formula, highlighting the corresponding place on the graphical screen. See (Archambault et al., 2007) for a more thorough description of how the Braille version of the interface works technically.

Apart from the studies described above, which clearly show that haptic and audio feedback can be used to develop shared interfaces for collaboration between sighted and visually impaired users, a number of single user applications were also developed within MICOLE (Saarinen et al., 2005; Crossan and Brewster, 2008) and other projects (Magnusson et al., 2002; Yu and Brewster, 2003; Wall and Brewster, 2006). These studies are all representative for the thesis study presented here in that they aim at developing haptic and/or audio tools that support learning and that are dependent on active exploration of different kinds of content. Saarinen et al. (2005) developed a software agent for multimodal input/output for use by visually impaired pupils. It is outside the scope of the thesis to describe the architecture here, but the interested reader should consult the paper. Saarinen et al. (2005) used their architecture to develop a learning system for exploration of different phenomena in the solar system. The simulation consists of six micro-worlds and a central station and the user can go from the central station to a micro world by opening doors. Each micro-world has a different focus (earth orbit, atmosphere, etc.) and can be explored by listening to verbal descriptions and by exploring interface elements haptically.

Crossan and Brewster (2008) developed and evaluated a system for haptic and visual trajectory playback of simple shapes, which resembles the work described earlier by Rassmus-Gröhn et al. (2007). Two experiments were carried out in which the visually impaired user should recreate shapes played back by the system. In the first experiment the user was haptically guided along the edges of a simple shape, through a pre-defined path, after which the user should recreate the shape felt. The shape was re-created by pushing the button on the haptic device handle while moving the proxy along the path felt during the playback. When performance was compared between sighted and visually impaired persons it was found that visually impaired pupils generally found it harder to recreate shapes felt. As a response to this Crossan and Brewster (2008) conducted a second experiment where shapes were played back by both haptic and audio feedback.
Different audio cues were used to present horizontal and vertical position of the proxy and the respective endpoints of the trajectory. A haptic only condition and a haptic/audio condition were compared in the experiment which showed that the recognition of the shape was significantly better when the additional audio cues were used.

Brewster has also been involved in studies investigating visually impaired users’ exploration of bar charts (Yu and Brewster, 2003; Wall and Brewster, 2006). As did (Crossan and Brewster, 2008) these studies also highlighted the importance of providing additional audio feedback to aid visually impaired persons’ exploration. Yu and Brewster (2003) performed several experiments related to bar chart exploration by means of haptic and auditory feedback. They developed bars that could be felt haptically and a midi tone whose pitch revealed the height of the bar the user was currently exploring. In an evaluation the participants solved tasks related to general trends and approximate values and results showed that tasks could be solved efficiently when using the one-point interaction PHANTOM device. More importantly, the results also showed that the tasks of reading values and spotting general trends were solved much faster when audio feedback was provided. Again, this indicates that there is a potential of adding sound to a haptic environment in interfaces for visually impaired computer users.

As this section shows, haptic and audio feedback can be used to successfully support blind or visually impaired computer users’ interaction in virtual environments. Before the MICOLE project there were very few projects that aimed at using haptic feedback to specifically support collaborative task solving between sighted and visually impaired pupils in a classroom setting. The applications developed within the scope of MICOLE which were discussed in this section, show great promise, but still lack task solving on equal terms, enabling joint handling of objects. The haptic applications described in the next section, which were developed within the scope of this thesis, utilizes functions for haptic real-time communication which has been lacking in earlier applications in this field of research.

4.3 Developing new interfaces

Three new applications were developed within this thesis project. The overall goal with the applications was to present new means of reaching common ground between sighted and visually impaired pupils while collaborating in geometrical task solving. A lack of common ground is a big cause of the problems discussed earlier in this chapter and if the developed applications make it possible to collaborate and communicate about joint task material, strong arguments about inclusion of visually impaired pupils can be made.

All applications were based on the Reachin API – a commercial API for programming haptic interfaces. Even though the commercial nature of the software was troublesome, this API was chosen since it was used in all the applications evaluated by Sallnäs (2004). Reachin API is based on the scene graph principle in that it consists of nodes and attributes which are combined together to form a complete 3D scene. Simple applications which do not contain any dynamic behavior can be developed by just nesting built-in graphic and haptic nodes with different attributes in a vrml-file. More complex applications, like the ones evaluated in the thesis project, require invoking python-scripts
for event handling and creation of new haptic nodes in C++. More details will be provided when necessary, in the presentation of the developed applications.

In this section the applications and the design process will be described. For convenience they will be referred to as the static and dynamic applications from here on. This naming has also been used in the main publication for this study (Moll and Sallnäs, 2013). Why the distinction between static and dynamic is important will become apparent in the result section later on.

Figure 8. The static application, for illustrating angles and geometrical figures. Taken from (Moll and Sallnäs, 2013).

Figure 9. The dynamic application for illustrating geometrical concepts like 3D shape and area.
Figure 8 and 9 above shows representative screen shots of the two applications used in the evaluations presented later. Both these applications were developed based on ongoing discussions with elementary school teachers to make sure that mathematical phenomena relevant to the pupils’ current state of knowledge would be presented.

4.3.1 The static application

The static application (figure 8) was designed to illustrate 2D geometric figures and angles. The reason why this application will be called “the static application” from here on is that all interface elements are stable during the interaction. This proved to be of utmost importance for both awareness and common ground, as will be discussed later on in the result section.

The basic interface consists of a flat surface (referred to as “a shallow tray” in (Moll and Sallnäs, 2013)) surrounded by walls to prevent visually impaired users from accidently leaving the surface and to clearly frame the working area. The walls are also meant to serve as reference points, which are of great importance when visually impaired persons explore an environment by touch (Sjöström, 2002). These interface elements are created in a simple VRML file by nesting simple Box nodes in the ReaChin API. The three buttons positioned under the main working area are used by the teacher to create new lines and to load and save tasks (consisting of a set of drawn figures). These buttons can easily be disabled from the VRML file so that no accidental events occur when visually impaired users explore the interface.

The events generated when pushing the buttons are carried out in a separate python script. Every time the “New line” button is pushed a new pre-defined line (a horizontal 2mm thin box) appears in the center. The preferred size and shape are selected with the Phantom from slide bars and the line can also be picked up and moved to the intended location. The solutions for adding new objects to the scene and moving the objects are taken from a collaborative design application evaluated by Sallnäs (2004). In figure 8 above an example is shown where lines have been combined to form four different basic geometrical shapes.

Since all lines are basic Box-nodes which have built-in visual and haptic attributes, the lines and their intersections can be felt. This makes it possible to explore all shapes drawn on the flat surface by following their contours. The friction on the lines is set to be slightly higher than on the flat surface and the walls to make it easy to distinguish between basic interface elements and parts of drawn shapes.

The static application, as well as the dynamic one, is meant to be used by two users working together to explore and/or categorize drawn shapes. Thus, the application works with two (or more) Phantoms. However, the static application does not include any functions for dSSirect haptic communication between the hardware devices as the ones discussed in the theoretical background section 2.3.2.
4.3.2 The dynamic application

The dynamic application (figure 9) was designed to illustrate 3D shapes and the concepts of area and volume. The term dynamic refers to the fact that the scene is constantly changing due to the users moving different objects around.

In this case two versions of the application were created – one for the teacher and one for the pupils (the one tested in the evaluation). Figure 10 shows a screen shot of the teacher version, where tasks are created.

![Figure 10. The teacher’s version of the dynamic application](image)

In the teacher’s program version new tasks are created by adding new basic objects to the scene. The objects to choose from are selected from the box in the upper left corner by pressing against the cylinder, cone, cube or sphere respectively with the haptic device. After selection, values relevant for the particular object can be chosen on haptic slide bars (e.g. height and radius for the cylinder and x-, y- and z-sizes for the cube). Stiffness and friction can also be set as well as the color. When all objects of choice have been placed in the scene the object positions and configurations can be saved to a simple text-file. Figure 10 shows an example where a box has been created after which it has been colored green and been given new dimensions. The methods for creating objects, changing values and saving files are the same as those for the static application. Since the teacher’s version of the dynamic application was never evaluated in a formal study it will not be described further in this thesis. It is not described in any of the articles or conference papers related to this study either. Note, however, that all interface elements and functions described later on in this section are present in both the teacher’s and the pupils’ version.
In the dynamic application the basic environment is a ceiling-less room seen from above, as in figures 9 and 10, with a box positioned in the upper left corner, containing an infinite set of the shapes cylinder, cone, cube and sphere. The texture applied to the floor, walls and box respectively can be felt and easily discriminated from each other. The textures are all based on gray scale bump map replicas of the images seen in the graphical interface. The height of the highest part of the maps is set to 0.2 mm to prevent especially the visually impaired user from being hindered in their exploration of the environment. The reason for applying discriminable textures to all basic interface elements is to assist visually impaired users in their exploration and navigation in the environment. By pushing the load button located outside the main working area a task created in the teacher’s program version can be loaded. In figure 9 above a task involving 10 cubes, placed between the box and the right wall, has been loaded.

The most interesting parts of the application, when it comes to the main topics of this thesis, are the haptic functions that enable interaction with the objects and between the users. Most of these functions have been inherited from the application evaluated in the hand-off study by Sallnäs et al. (2000) described earlier. The functions and their application in this part of the thesis work will be explained next.

As for the objects they can all be “grabbed” by pushing a button on the haptic device when the proxy is in contact with them. At the moment when an object is grabbed an invisible virtual band is created between the centers of the proxy and the object respectively. Since elasticity of the band is simulated and its stiffness is high the proxy will be drawn towards the object when the button is pushed. In this way, the user perceives that he/she is grabbing the object directly. A general force field is also applied which makes it possible to experience the weight of the object as long as it is held (as long as the button is pushed). The elastic forces from the band also make it possible to swing objects around and to feel torque and inertia. As stated before, the virtual band is not perceived by the user. If two users grab the same object two virtual bands are created as described above. The users can then feel each other’s forces on the jointly held object through the respective bands. This gives the users a very realistic feeling of joint manipulation. Since this function can be used to guide e.g. a visually impaired user the function will be referred to as a “guiding function” in the result and discussion sections.

The last haptic function enables grabbing the other user’s proxy directly. When the two proxies are in contact and one user pushes the button on the haptic device a virtual magnet is simulated in the center of the proxy corresponding to that haptic device. As long as the button is pressed the two users are then “glued together” due to the magnetic force generated from one of the proxies.

4.4 Evaluation in the field

The method chosen for this particular study was an informal field evaluation and thus the results were mainly based on observations and interviews with pupils. The reason for this choice of method was mainly that we were to be working with disabled pupils in elementary school. Apart from the fact that severe visual disabilities can take many forms, different visually impaired pupils are also accustomed to different work material. Thus, a qualitative approach less sensitive to individual differences was the preferred one in this
case. The evaluation carried out and the choices made will be more thoroughly described in the remainder of this section. For even more details about the evaluation, especially with regard to working with kids, the article (Moll and Sallnäs, 2013) should be consulted.

4.4.1 Subjects

In this study the choice was made to involve pupils from the target user group – visually impaired pupils. This choice had its pros and cons. On the one hand the results gained had a high validity, but on the other hand this choice limited the number of available groups to test with.

In total, four groups of pupils participated in the evaluation. Each group consisted of three pupils of whom one was severely visually impaired or blind. The ambition from the beginning was to let three pupils interact in the environment at the same time, but unfortunately only two hardware devices could be connected to the computer, which meant that the sighted pupils needed to take turns in using the haptic equipment. The pupils in the respective groups went to the same class and hence knew each other in advance and were used to working together. The groups came from four different elementary schools in the Stockholm area.

4.4.2 Apparatus

The computer used in the evaluation was a Dell Precision PC with two dual-core processors and the necessary applications ReachIn API 4.1 and Visual Studio .NET 2003 installed. A flat LCD-screen, for the sighted pupils to look at, was connected to the computer as well as ordinary keyboard and mouse, used by the researcher for loading the different tasks.

Due to hardware limitations two PHANTOM Desktops could not be connected to the computer and thus one PHANTOM Desktop and one PHANTOM Omni (from hereon called just Omni and Desktop) were the devices used in the evaluation. These two devices are similar in appearance and usage, but are different when it comes to resolution – the Omni has notably less resolution than the Desktop. The consequence of this is that the user controlling the Omni has a harder time feeling e.g. texture details. During the evaluations the sighted pupils got to use the Omni since they also had the graphical interface to support them in their interaction with the interface. The differences between the two devices most probably did not affect the task solving process.

4.4.3 Setting

The evaluations were carried out on-site in the schools in which the subjects studied. Figure 11 shows a screenshot of the setting used on all evaluation sites. All four evaluations were performed in a room close to the classroom in which the pupils normally studied, so that the surroundings were known at the same time as disturbances
were kept to a minimum. All the necessary equipment was brought to the school on the date of the respective evaluation.

As figure 11 illustrates, the pupils sat close to each other since this is the setting normally used in group work situations in school. The sighted pupils sat beside each other taking turn in using the Omni and the visually impaired pupil always used the Desktop. The flat screen was positioned on the table so that it was easily visible by both sighted pupils. Since the screen was lying flat on the table they saw the scene from above (through the invisible roof). This made it easier for the pupils to build a common frame of reference regarding direction words, since the scene is also felt from above by both the sighted and the visually impaired pupil.

In the upper right corner of figure 11 one can also see the visually impaired pupil’s personal assistant. In three of the groups the personal assistant was always close by. The choice of having the assistant in the same room was the visually impaired pupil’s alone.

4.4.4 Procedure and tasks

All in all, the evaluation consisted of five parts in which the pupils were active; (1) training in the static application, (2) task solving in the static application, (3) training in the dynamic application, (4) task solving in the dynamic application and (5) concluding interview. These steps were always carried out in the same order.

After introductory information about the purpose of the evaluation and a short description of the outline of the entire session the first training task was loaded in the static application. The task that the pupils were presented with is shown in figure 9 above. The focus of this first training task was to get the pupils acquainted with using the haptic feedback and to show them how haptic feedback can be used for recognition of simple shapes. After the visually impaired pupil had been manually guided to the flat surface he/she and the sighted pupils were asked to go around the workspace following the edges, in order to get a feeling for the workspace dimensions. After this the pupils were asked to collaborate in finding and naming the four geometrical shapes drawn on the flat surface.
The important parts here were to train the pupils in navigating between different shapes as well as showing the pupils that it is possible to feel specific shapes with the haptic equipment.

After the training task, which lasted about 10 minutes, the first test task was loaded in the static application. Figure 12 below shows the task.

![Figure 12. The test task in the static application.](image)

The pupils were then given the following directions, verbally (the formulation “the box” was used to describe the flat surface with surrounding walls to the visually impaired pupil):

*Task 1: at the bottom of the box there is a new, large shape. You shall now decide together for each angle in the shape whether it is right, obtuse or acute.*

When all angles had been correctly classified and all pupils agreed on the solution the static application was closed and the training task in the dynamic environment was loaded. The scene directly after startup is shown in figure 13 below. The focus of this training task was to get the students acquainted with the special haptic functions described in section 4.3.2 about the dynamic application.

After locating the walls, corners and the box the pupils were asked to locate the set of cubes positioned between the box and the room’s right wall. The procedures for grasping, dropping and handing off cubes were then explained and the pupils practiced a few minutes until they were fluent in using these functions. Last, the procedures for grasping the same object and each other’s proxies were explained and these were also practiced. The training continued until it was apparent that the pupils understood how to use the specific guiding functions and could apply the techniques effortlessly. This was to make sure that the group work during the test task would not be hindered by difficulties in handling the technology.

After the training task was finished the same scene was loaded again, but this time as a test task. The following directions were given to the pupils:
Task 2: there are, as before, ten cubes to the right of the box. All cubes have sides of 2 cm. You shall now work together and place cubes on the floor so that they cover a rectangle with an area of 32 cm². Tip: How large an area does one cube cover?

Before the pupils started working with the assignment it was emphasized that they should communicate and collaborate during the task solving process. Figure 13 below shows an example of how the workspace looked when the pupils in one group had been working for a while.

![Image](image.png)

Figure 13. The pupils are working with the test task in the dynamic application. In the middle of the scene they can be seen carrying the same cube.

When all pupils were in consensus that the task had been solved the application was closed.

During the last 10-15 minutes of each evaluation session the pupils were interviewed in groups all three together. The questions concerned usability issues, general attitudes towards the application, the level of inclusion and the potential for learning.

4.5 Analyses

The qualitative analyses were based on video recordings and interviews. Both these data sources were analyzed using a content analysis approach, resulting in emergent themes of special interest. The general procedure followed is explained in section 3.1 about analyzing dialogues. The categories relevant for this thesis are listed in the next two subsections.

The interviews were also transcribed in their entirety. After transcription, the different parts of the interview were summarized and sorted under the respective questions asked. This was done for each group. Answers from the groups were compared in a last iteration.

At the outset of the project presented in this chapter the only aim was to collect qualitative data about several aspects related to CSCW and CSCL (described in detail below). As the analysis was iterated in response to later findings and article review processes, however, several interesting quantitative measures also arose. In this section both the quantitative and qualitative data used in the last iteration will be described.
4.5.1 Quantitative measures

- **Time to perform task**
  The time it took for the pupils to solve the respective test task. The time started when the respective task was loaded and stopped when all pupils agreed that the task was solved. This data was mostly used as a performance measure and to compare the different groups. The unit was whole minutes.

- **Time taken for verbal guiding operations (both group wise and combined)**
  The time it took, in seconds, to perform a verbal guiding operation. Both mean and average time was calculated for each group and all groups combined, respectively. The total time spent on verbal guiding and the number of verbal guiding operations were also calculated. A verbal guiding operation is in this case defined as “a sentence or other sequence of words explicitly leading the visually impaired pupil to a certain place” (Moll and Sallnäs, 2013). The time is taken from the moment the verbal guiding starts until the visually impaired pupil has reached the intended target place. This measure is particularly interesting when compared with time taken to perform a haptic guiding operation.

- **Time taken for haptic guiding operations (both group wise and combined)**
  The time it took, in seconds, to perform a haptic guiding operation. Both mean and average time was calculated in the same way as for verbal guiding operations. The total time spent on haptic guiding and the number of haptic guiding operations were also calculated. In this case, a haptic guiding operation is defined as “taking place as long as two people are grasping the same cube or each other’s proxies, approaching a pre-defined goal” (Moll and Sallnäs, 2013).

4.5.2 Qualitative measures

- **Common ground**
  Here, Clark and Brennan’s (1991) definition is used. Means of achieving common ground, by means of verbal and haptic communication, regarding the task, the workspace layout and the objects present are considered here. For this study the most important issue considered is how haptic feedback directly influences the dialogue and the reasoning about the objects and what is being done.

- **Awareness**
  Here, Dourish and Bellotti’s (1992) definition is used. The focus here is on the visually impaired pupil’s understanding of what is currently going on and knowledge of the sighted pupil’s position in the interface. For this study the most important issue regarding awareness is how the haptic feedback can be used to provide useful awareness information to the visually impaired pupil.
**Guiding**

Verbal and haptic guiding are defined in 4.5.1 above. Here, results regarding the means of providing and the reasons for providing verbal and haptic guiding are considered. This is the most important qualitative data collected when this thesis is concerned, since it sheds light on how and why haptic communication is used.

**Navigation**

Here, the use of haptic feedback to navigate in the virtual environments is considered. The results on this qualitative measure will only be discussed briefly, since it does not relate to the main focus of the thesis. Note that this measure concerned the visually impaired pupil’s own navigation. Results on joint navigation, i.e. guiding, are covered by the Guiding category.

**Initiative**

Here, the means by which the sighted and visually impaired pupil, respectively, took and kept the initiative in the task solving process are considered. As for navigation, results on this qualitative measure will only be treated briefly due to its limited relevance to the main focus of the thesis.

### 4.6 Main results

In this section the main results gained from the study will be provided and discussed. Dialogue excerpts will be provided to illustrate the main findings. In these excerpts the visually impaired pupil is denoted V and the sighted ones S1 and S2. For more examples and a more thorough treatment of some of the qualitative data listed above the reader should consult the article (Moll and Sallnäs, 2013).

All in all, the pupils managed to solve both tasks and table 1 below summarizes the time taken for each group.

<table>
<thead>
<tr>
<th>Task</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Task 2</td>
<td>15</td>
<td>12</td>
<td>11</td>
<td>16</td>
</tr>
</tbody>
</table>

In general, the visually impaired pupils experienced no problem in controlling the haptic device or in interpreting the haptic feedback. In the static application the visually impaired pupils could accurately categorize all angles and figures and in the dynamic application they could easily apply all haptic functions – they quickly learned how to grasp cubes, drop them and carry them together with a sighted peer.

In the following result sections some special themes, directly related to the influence of haptic feedback on communication, will be elaborated on.
4.6.1  **Haptic feedback’s influence on verbal communication**

One of the main themes in this thesis concerns how haptic feedback influences the dialogue in collaborative virtual environments. This potential influence can affect the entire collaborative activity and is thus important to investigate. In this regard this first study provided a number of interesting insights.

Means of establishing common ground were clearly influenced by the haptic feedback. Example 1 below, taken from the first task, illustrates a typical phenomenon. Before the dialogue starts the visually impaired pupil had just found the trapezoid shape, which they had never talked about before in class.

**Example 1:**

V:  *This is a sort of rectangle, right?*
S1:  *No, not quite so…*
V:  *What is this, then?*
S1:  *It is…, we have never talked about it…*
V:  *I do not know its name, we have not read about it*
S1:  *No, I do not know…*
S2:  *hm, …, maybe, …, no…*
V:  *No, I do not know its name, but I have heard about it before.*
   [Now, we tell them that it is a trapezoid]

In the above example there is no doubt about that all pupils share a common ground about the object in focus. It is also clear that the visually impaired pupil is aware of the fact that both proxies can be seen in the interface and that they are all talking about the same object. These are all important prerequisites for effective and meaningful collaboration and communication. In example 1 it is particularly interesting to look at the means of referring to the object as “this” and “it”. No attempts are made in trying to define the object’s position in relation to the other objects or workspace boundaries, which would have been necessary if only verbal description of the content seen by the sighted pupils was possible, as confirmed by Sallnäs et al. (2005). The means of deictic referencing shows that the pupils share sufficient common ground and this effortless discussion would never have taken place without the visually impaired’s access to haptic feedback. It is also clear from the above example that the common ground that the haptic feedback provides makes it possible for the pupils to focus on the actual task in their discussions and not on gaining consensus about the content of the workspace. The example also illustrates a common ground based on shared history – the pupils acknowledge the fact that they have not encountered the object before. The feedback gained from the haptic feedback makes this connection to the shared history possible. A last important point to be made with regard to example 1 is that it is easy for the visually impaired pupil to use the haptic feedback and to interpret what is felt.

Example 2 below provides some more examples of how the dialogue is influenced by the access to haptic feedback.
Example 2:

[V holds a cube]

V: Should I put this down here?

S1: Ok, wait... Follow me here...

[S1 starts to guide V haptically]

S1: Up, up... ok...

S2: down there... a little bit more

[S1 still provides haptic guiding]

S1: Ok, that’s it!

Although this is not very common, the sighted pupil in the above example provides complementing verbal guiding during the haptic guiding operation. The example starts out by the visually impaired pupil holding a cube which she just picked up. Then she refers to the cube as “this”. The fact that she feels that she is holding the cube and that she is aware of the fact that two users are present in the visual/haptic interface ensures her that this utterance is meaningful. The communicative properties that the haptic feedback functions have also connect the pupils together in that the participants can feel forces from each other in the environment. This is demonstrated when the sighted pupil says “follow me”. This is a kind of directional reference that depends on the pupils’ feeling physically connected via the haptic feedback provided.

A last important point, which is illustrated in both examples 1 and 2, is that the visually impaired pupils have the initiative. In example 1 the visually impaired pupil is clearly leading the discussion. The shared view made possible by the haptic feedback enables the visually impaired pupil to take this initiative, by always being the one pushing forward in the conversation about the unknown object. Even though the visually impaired pupil is guided by one of the sighted pupils in the second example, the visually impaired pupil still has the initiative and is also the one requesting guidance. Both examples show that the visually impaired pupil can take the initiative and as a consequence he/she can drive the discussion forward. The haptic feedback provided makes this possible.

4.6.2 Providing guidance

Another of the most important results gained from this study concerned the reasons why haptic guiding was used and the effects of using it. Since the groups were few, for reasons explained earlier, the quantitative results regarding guiding cannot be used for statistical analysis. However, these results nevertheless show interesting patterns. Table 2 below shows how much time the different groups, on average, spent on one verbal or haptic guiding operation respectively. Since the mean and median times were very similar only the mean times are provided in the table. Also note that haptic guiding could only be used in the dynamic application (task 2). In the fourth group, G4, the pupils never made use of haptic guiding.
Table 2. Quantitative results regarding verbal and haptic guiding. Means for one single verbal and haptic guiding operation respectively are provided (values in red). All values in seconds.

<table>
<thead>
<tr>
<th>Group</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G1-G4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>10.5</td>
<td>21.0</td>
<td>10.4</td>
<td>15.6</td>
<td>16.2</td>
</tr>
<tr>
<td>Haptic</td>
<td>5.0</td>
<td>10.6</td>
<td>4.56</td>
<td>xxx</td>
<td>6.29</td>
</tr>
</tbody>
</table>

Figure 14 and figure 15 below presents the rest of the quantitative data on guiding operations – number of guiding operations per task and the total times spent on providing guiding in each task.

Figure 14. Number of guiding operations per task and group respectively. Taken from (Moll and Sallnäs, 2013).

Figure 15. Total time spent, in seconds, during the different guiding operations for each group and task. Taken from (Moll and Sallnäs, 2013).
The values in table 2 show that the average time spent on haptic guiding is much shorter than that spent on verbal guiding. Since verbal and haptic guiding are two means of establishing the same type of goal – leading the visually impaired pupil to a certain place – the big difference between the times can hardly be explained by the haptic guiding operations always being simpler or concerning a shorter distance. Hence, it is quite clear from the quantitative data in table 2 that haptic guiding is more efficient than the verbal counterpart. These results are also supported by figures 14 and 15 which show that considerably more time is spent on verbal guiding compared to haptic guiding, despite the fact that the number of haptic guiding operations is actually larger for one of the groups.

These results could in large be explained by the qualitative results regarding guiding. Example 3 below, taken from the second task, illustrates the main points regarding verbal and haptic guiding. The example is illustrative in that several similar excerpts were found throughout the transcripts.

Example 3:

S1: All right, you can pick that one you have now
[V picks up the cube]
S1: And then, ..., wait, ..., a little bit more to the right. Up, take it up a little, ..., [V moves up towards the roof]
S2: Here, here, towards you.
[S2 corrects the direction cue…]
[V moves down again and reaches the cube]
S1: Good!
V: No, it does not work
[S1 picks up the same cube to help]
[They move the cube towards them]
[They move the cube a tiny bit to the left]
[They place the cube to the right of another one]
[They fine-tune the cube’s position]
S1: That is good!
[V and S1 let go of the cube]

The example starts out with the visually impaired pupil picking up a cube, which he is currently close to. The rest of the first half of the example illustrates why verbal guiding could be problematic. It is clear that quite long sequences of words are needed to guide the visually impaired pupil to an intended spot. Guiding the visually impaired pupil, step by step, in this way is cumbersome and after a while the visually impaired pupil gives up. The following utterance provides yet another example of the limitations of verbal guiding.

“go down..., slowly, slowly..., left..., right, then up... there you go!”

Obviously it is quite hard and cumbersome to give verbal directions. In the second part of example 3 haptic guiding is being utilized and yet again interesting patterns arise. First of all nothing is said during the haptic guiding operation indicating that the haptic feedback
in itself provides enough cues about the ongoing activity. Second, the visually impaired pupil keeps holding on to the cube and hence follows along the sighted pupil’s movements. Even though the visually impaired pupil is not the leading force in this example, the haptic guiding function (holding on to the same cube in this case) in this way provides important awareness information to the visually impaired pupil. Unless the sighted pupils constantly verbalize the changes made and the steps taken towards the final solution, this type of awareness information, about actions taken by the sighted peer, would be inaccessible to the visually impaired pupil. This, of course, is detrimental when it comes to inclusion of visually impaired pupils in group work with sighted peers. As an interesting side note it is also worth mentioning that the first part of the above example, up to the point where the haptic guiding started, took about 23 seconds. When the pupils found out that it was too cumbersome to give verbal directions they switched to haptic guiding and the placement of the cube was completed in 5 seconds. This might be the most important result when it comes to guiding in this study – the pupils start out using verbal directions as they are used to and switch to haptic guiding when they realize that the verbal alternative becomes too cumbersome.

Example 3 above also illustrates how the haptic guiding function influences the dialogue. As already stated the example started out with the sighted pupils providing verbal directions and ended with one of the sighted pupils giving haptic guiding. The fact that nothing was said during the haptic guiding procedure indicates that haptic feedback replaces the need for verbal directions. This illustrates that some things, like directions and intentions (follow along or resist) can be communicated by the haptic feedback alone. This results in an important gain, since the example shows that verbal guiding could be time consuming and cumbersome. During a haptic guiding operation enough information is provided from the haptic feedback itself – the visually impaired pupil feels directions and the approximate length of movements. Several other examples on how haptic guiding was used by the pupils, and how it influenced the dialogues, can be found in the article (Moll and Sallnäs, 2013).

4.6.3 The static vs. dynamic environment

An important issue raised during the evaluation was the lack of awareness of changes that the visually impaired pupil had access to in the dynamic application. During the analyses it became apparent that the most important reason for this was the dynamic nature of that application – hence the application was named “dynamic application”. Even though it was not a focus at the outset of the study, the distinction between static and dynamic turned out to be of utmost importance.

In the dynamic application the situation was problematic mainly because the objects were constantly moved by the sighted pupils in the interface. When the second test task was started the visually impaired pupils soon started to search for the cubes between the box and right wall – a common strategy for achieving workspace awareness. The problem is that not only walls and corners but also the cubes themselves were used as reference points for navigation in the same way as workspace boundaries and parts of figures were used as reference points in the static application. When the cubes were moved, especially
by the sighted pupil, the visually impaired one often got disoriented and lost the orientation. This was often illustrated by utterances like;

“Where did that one go?“ [when feeling absence of the cube believed to be there]

The haptic guiding function decreased this problem to a large extent since both of the pupils had control of the jointly held cube and could feel the direction and length of the movement. In several cases, however, the sighted pupils moved cubes on their own and in these cases there was no way for the visually impaired pupil to know that a cube had been moved and to what location. The only thing the visually impaired pupil may notice is that a cube has been moved from its original location.

Apart from minor usability issues the pupils experienced no problem in working and collaborating in the static environment. One reason is that the environment was stable in nature since nothing could be moved. Once the visually impaired pupils had explored the interface for a few minutes they knew exactly where the different shapes or angles were. Example 4 below presents a typical case during exploration of the star-shaped figure in the first test task. Figure 16 shows the star-shaped figure with numbered angles. These numbers are referenced, with [x], in example 4.

![Figure 16. The star-shaped figure, explored in the static application, with numbered angles.](image)

Example 4:

**V:** I think this [1] is a right angle  
**S1:** Yes  
**V:** And this one [2], then. Is it acute?  
**S1:** It’s acute  
[Now, the visually impaired starts to recapitulate what has been said]  
**V:** This, this [3] was obtuse, right? ... And this one [4] acute...  
[She visits the last two angles once again]  
This [3] is obtuse, this [4] is acute.  
[She now moves to the ones just discussed]  
This [8] is obtuse...
The example shows that the visually impaired pupil has no trouble navigating in the star-shaped figure. She knows her way around the figure and can even tell the type of angle before getting to it. The example also shows that the visually impaired pupil is truly included in the work and that there are no problems in achieving common ground on the basis of what is felt and seen respectively. In the static environment the visually impaired pupil often started out in control and continued being the leading force during the entire task solving process. As can be seen in example 4, the visually impaired pupil led the discussion and decided what part of the figure to focus on. The static application provides an inclusive shared working area in which collaboration between visually impaired and sighted pupils can take place almost without problem on either side. This is much due to its static nature. The technology in itself is a prerequisite here – the haptic feedback experienced through the haptic device provides the visually impaired pupil with the information necessary to lead the collaborative activity.

Even though the visually impaired pupils sometimes managed to take control in the dynamic application as well the problems with awareness and common ground in that application were apparent. There were clearly important implications for design here – in the dynamic application the haptic feedback provided was not enough. There was nothing wrong with the haptic feedback in itself but rather more information was needed than could be provided through the haptic sense.

4.7 Lessons learned

Even though there was a problem with awareness in the dynamic application the evaluation showed that the applications used did include the visually impaired pupil and that especially the haptic functions were an aid in the collaboration. In this section the most important themes and lessons learned will be discussed more deeply.

4.7.1 Communicating with haptic feedback

The most important results of the study, for this thesis, were the ones concerning how the haptic feedback influenced the collaboration between the visually impaired and sighted pupils. At the outset of this study the focus of the work was mainly on how haptic feedback could be used to ease inclusion of the visually impaired pupil in group work. This was done by considering the common ground reached and how the visually impaired pupil got awareness of the ongoing activities. No attempts were made to look at how haptic feedback in itself can be used for communication or how it influenced the dialogue. After a few iterations of the transcripts, however, the potential that haptic feedback has when it comes to communication were made clear. Based on the results from this study
“haptic communication” became one of the key focuses in the research and the results from this study laid the ground work for big parts of the research presented in this thesis.

As discussed in the theoretical section haptic feedback has been used for collaboration before, involving sighted and visually impaired users (Rassmus-Gröhn et al., 2007; Plimmer et al., 2008), but these studies differ from the current study in a number of respects, with regard to how the haptic feedback was investigated and interpreted. In the study by Rassmus-Gröhn et al. (2007) they focused on 2D interaction. In that way there was a similarity with the static application used in the thesis study. Their focus, however, was not on how the haptic feedback influenced collaboration and communication but rather on how the application may be used as a presentation medium and how access to the shared interface affects the inclusion of the visually impaired pupil. The guiding functions used by Rassmus-Gröhn et al. (2007) differs from the ones in the thesis study in that only the visually impaired pupil got the haptic feedback. In this way the communication enabled by the guiding function is not reciprocal in that both users do not feel the feedback – the sighted user cannot feel if the visually impaired one wants to follow along or not. Additionally, the “assistant” guided the visually impaired pupil and not the other pupil. Thus, one of the prerequisites in that study was that the visually impaired pupil should get assistance from a third person while interacting with the system under evaluation. In the thesis study the dynamic application enables joint manipulation felt by both users and in that way it is more natural to do things together and it is also simpler to convey subtle intentions during the interaction.

In the study by Plimmer et al. (2008) introduced earlier, a kind of constant force guiding was used to transfer a teacher’s hand movements to a visually impaired pupil. Thus, the pupil is forced along a pre-defined path, decided by the teacher, and has no means of deviating from the path. Although some sort of one-way information transfer exists between the teacher and the pupil the communication made possible by the haptic channel is more restricted than in the dynamic application discussed in this chapter. However, both studies, along with (Rassmus-Gröhn et al., 2007) show potential of haptic feedback when it comes to adding an extra channel for communication.

It is interesting to view these results in light of the results from the field study reported by Sallnäs et al. (2005). A main problem identified there was that pupils experienced problems in sharing the work material. From the result of the thesis study it becomes quite clear that haptic feedback can provide a shared workspace to which both visually impaired and sighted pupils can refer. Especially in the static application the haptic feedback made it possible to communicate effortlessly and share the common workspace in a natural way. From the observations made it was clear that the visually impaired pupils were fully involved in the work and that they often took the initiative as some of the provided dialogue examples show. Thus, the evaluation clearly shows that one can use haptic feedback to create a shared interface which both sighted and visually impaired pupils can refer to easily. This finding is especially interesting to relate to the results gained by Storch (2002) when observing different roles in group work. In the field study within the MICOLE project the pattern “dominant/passive” often arose in groups including visually impaired pupils (Sallnäs et al., 2005). Especially in the static prototype, discussed in this chapter, it was clear that the pattern “collaborative” or maybe in some cases “dominant/dominant” arose. There were very few cases in which the visually impaired pupil took a passive role and in light of earlier discussions this is good for the
group work process. The haptic feedback enabled the construction of a joint problem space and a shared frame of reference. This in turn supported the communication between the participants – a communication that all participants took active part in.

That the common frame of reference provided by the haptic feedback, in both applications evaluated, influenced the communication between the participants is clear from the results of the study. Even though there are no functions implemented for direct communication between the hardware devices in the static application it is still possible to see how the haptic feedback can be used for deictic referencing. As was shown in several examples in section 4.6, the objects felt are referred to as “this”, “that”, etc. Without the common ground shared by means of the mutual access to the same material such effortless discussions about the objects would not be possible to carry out. The haptic feedback in itself could not be used to provide deictic references, since the users could not feel each other’s’ forces in the static environment, but it laid the ground work for effective verbal referencing.

In the dynamic environment the influences that haptic feedback had on the communication was even more explicit. As illustrated in section 4.6 haptic guiding was used several times to ease communication. It cannot be stressed enough that giving detailed verbal directions, especially to a visually impaired person moving in a 3D environment (whether it is real or virtual), is cumbersome and time consuming. The haptic guiding function clearly reduced the amount of verbal references needed. An analogy with learning to play golf could be made here. Most people know how hard it is to learn the right technique for making an appropriate golf swing. Verbally explaining how to swing is very difficult since posture, velocity, angles, etc. need to be perfect. The only practical way of explaining how to make a good golf swing is to grab the golf club from behind the novice player and follow him/her through the movement.

By using a haptic function, through which both users can feel and provide feedback continuously and independently at the same time, information about direction and the intention to follow along can be conveyed without any explicit verbal references. In light of this it seems natural to introduce the term *haptic deixis*, to describe the intention of moving in a certain direction, conveyed through the haptic guiding function. In basic communication theory a deictic reference refers to either descriptive words or pointing gestures (Clark, 2003). The purpose of using a deictic reference in interaction with peers is to direct the partners’ attention to a specific object or a location. When the haptic guiding function, utilized in the evaluated dynamic application, is used this type of information is provided haptically. The reader should bear in mind that the concept of guiding is more complex than the concept of deixis – the important conclusion made here is that deictic references are subtly fed through during the guiding operation and that this is one of the reasons why the haptic guiding function is powerful when it comes to adding a new element to the communication between the users. This discussion can also be related to Clark’s (2003) definition of placing. Since placing is a kind of indicative gesture moving an object into the addressee’s focus of attention, one could argue that haptic guiding through a joint object is a kind of joint placing action. This is especially true when both participants are moving an object into a certain position without saying anything. While holding on to the same object both participants are, by means of forces, making sure that they both have the same object in focus and they also share a common ground about the meaning of this action.

73
Apart from supporting communication in the shared interface the haptic guiding function (enabling manipulation of a joint object) enables a kind of haptic direct manipulation, according to the definition of direct manipulation given by Shneiderman (1997). Using a kind of drag-and-drop metaphor objects are moved in real-time, changing the configuration of the shared interface. Thus, at the same time as the guiding function mediates reciprocal communication between two users, it also enables joint direct manipulation. Nevertheless, haptic guiding is much more than direct manipulation of objects.

As stated earlier, the guiding functions used to evaluate haptic communication in this study have been tested before (Sallnäs et al., 2006). In that study they also showed that haptic feedback is highly important for forming a shared understanding about the layout of the scene, used to coordinate joint activities. They also touch on the subject of verbal and haptic guiding behaviors, but do not provide a detailed analysis of how access to haptic feedback influences the dialogue between the participants and the overall collaboration. This is where one of the main contributions from this thesis study can be found.

4.7.2 The importance of awareness feedback

One of the main rationales behind the study presented in this chapter was to study how haptic feedback in particular could ease the inclusion of the visually impaired pupils in the group work process. It is clear from the results that the construction of the joint problem space and the establishment of common ground made possible through the haptic feedback made a positive contribution to the inclusion of the visually impaired pupil in the static application. The guiding function used in the dynamic application was also shown to make it possible for the visually impaired and sighted pupils to do actions jointly and this is also positive for inclusion. One of the big problems, which could not be reduced by the haptic guiding function, was that the dynamic environment contained objects that could be moved by both the visually impaired and the sighted pupils. The reader may refer back to the discussion about the difference between the static and dynamic interfaces in section 4.6.3.

The core problem lies in the lack of awareness about changes made in the environment. Without knowing that something is moving and where changes occur the visually impaired pupil had difficulties making informed decisions about what to do. The formal definition of awareness was introduced in section 2.2.1.1; “an understanding of the activities of others, which provides a context for your own activity” (Dourish and Bellotti, 1992). It was quite clear from observations that the visually impaired pupils did not understand most of the activities that the sighted pupils performed on their own, as was illustrated by examples in section 4.6. Comments from visually impaired pupils also show that they did not understand what actions to take in some occasions – they did not have an accurate context for making informed decisions about their own actions. Thus, it is quite clear that the big problems encountered during the study are the result of lacking awareness information. This being said, it is important to bear in mind that the task itself was dynamic in nature. Other types of tasks could also have been designed which were not dependent on the placement of blocks in a particular pattern. However, one of the
core ideas with the study was to see how a haptic real-time guiding function could make it possible to handle objects jointly and enable joint navigation. Many hands-on tasks in education are dynamic in nature and this is another reason why one of the evaluated applications contained a dynamic interface.

It is clear that it is straightforward for the visually impaired pupils to use the haptic functions for manipulating objects, something that has been shown in earlier studies on the same theme as well (Sallnäs et al., 2006). If an understanding about the other person’s actions in the interface would have been possible to achieve it is reasonable to believe that the visually impaired pupil would be a lot more involved in the work, due to the increased possibility of making more meaningful actions. Using haptic feedback to intuitively convey the necessary information (awareness about all changes made) would be hard, unless the two hardware devices are connected, by the haptic guiding function, throughout the entire task solving process. Such a constant guiding, forcing both users to follow each other’s movements as long as the program is running, would include all users in the work process mostly because the visually impaired pupil is forced to follow the sighted’s movements. Awareness about actions taken would be provided through this haptic guiding function as described earlier. Although haptic feedback has been used in this way in studies referenced earlier (Rassmus-Gröhn et al., 2007; Plimmer et al., 2008), it would severely limit the visually impaired pupil’s possibilities to take initiatives. Again, we would get the “passive/dominant” state proposed by Storch (2002). One alternative could be to implement the kind of go-to function designed by Oakley et al. (2001). This would make it possible for the visually impaired user to track the sighted’s movements at his/her own will. However, this would still not provide the visually impaired user with the constant awareness update which is obviously needed.

The only choice left is to use the audio channel to convey the missing awareness information and several examples brought up in the theoretical sections has shown great promise with such an approach (Gaver et al., 1991; Hancock et al., 2005; Gutwin et al., 2011; Yu and Brewster, 2003; Rassmus-Gröhn et al., 2007; Magnusson et al., 2007). Using sound for conveying information about actions taken in the interface would potentially decrease the need for verbal guiding and let the pupils focus even more on the task at hand. Additional knowledge about movements and location of actions would also provide the visually impaired pupil with a more accurate working context from which it would be easier to make informed decisions. Due to the difficulties encountered in the evaluated interface and the potential of audio feedback to convey some of the missing information the decision was made to conduct the follow-up study discussed in the next chapter.
5  Haptic and audio feedback for collaboration between sighted and blindfolded adults

In this chapter I will present the second study involving collaboration between sighted persons and those who cannot see (in this case blindfolded). It was clear from the study presented in the last chapter that awareness problems arose especially in the dynamic application and that these had unfortunate impacts on the collaboration between the participants. In the follow-up study presented here the main aim was to investigate if added audio cues would solve the awareness problems and as a consequence facilitate the collaboration. Even though we used blindfolded persons instead of visually impaired ones in this study we were able to confirm the results on common ground and especially the influence of haptic guiding functions and the results also clearly showed that audio plays an important role for mediating awareness information.

This study was also carried out within the scope of the project MICOLE introduced earlier. Apart from me and Eva-Lotta Saltnäs Pysander, the research student Yingying Huang took active part in the study. After I had made necessary modifications to the dynamic application the three of us planned the entire study. Yingying and I ran most of the experiments together. The qualitative analysis was performed by me alone while Eva-Lotta and Yingying performed the quantitative analyses. The study resulted in two journal articles – one mostly about the quantitative results (Huang et al., 2012) and one about the qualitative results (Moll et al., 2010). I have written most of the qualitative result parts in both of these articles.

The test setting used in this study. A screen shot showing the current state of the interface is shown in the lower right corner.
5.1 Incorporating sound feedback in a haptic interface

In the study described in the previous chapter problems were identified which hindered the collaborative process when the pupils were collaborating in the dynamic application. Specifically, the visually impaired pupil had no means of acquiring awareness of changes made by the sighted peer. This is why the study presented here is based on a modified version of the dynamic application – a version that use sound as a means of presenting awareness information about the ongoing work.

As was discussed in the theoretical background sound has in many cases been successfully used to make users aware of the fact that a change has happened in the interface and of the cause of the change (Gaver et al., 1991; Gutwin et al., 2011). Gaver et al. (1991) also showed that a complex environment can be successfully described by a sonification running continuously in the background. Such sound models have also been used successfully to make it possible for visually impaired computer users to get a quick overview of a scene and to focus on specific parts of an interface (Winberg and Bowers, 2004; Winberg, 2006). In section 4.2 several examples were also given on how audio cues can give supporting information about the computer interface and the interaction (Magnusson et al., 2007; Rassmus-Gröhn et al., 2007; Crossan and Brewster, 2008). In their pioneering study Gaver et al. (1991) also specifically showed that the sound model implemented provided sufficient action awareness for joint task solving to be possible. Clearly, there is a potential in sound cues in collaborative virtual environments for increasing the amount of awareness information about the status of the interface and actions taken. In light of this, the dynamic application from the last study was updated to incorporate sound cues for actions taken.

In the version of the dynamic application tested in the study presented here four sound cues were used – three sound cues for interaction and one for locating the sighted user. Three separate sound icons were used for the interaction sounds, since the intention was to resemble real handling of objects (see (Gaver, 1993)). The first sound icon, which was used for the grasping event, was heard every time an object was grasped by either one of the users. From here on this sound will be referred to as the “grasp sound”. This sound icon plays a very short and distinct cymbal tone. The second sound icon, which was used for the dropping event, was heard every time an object was dropped on the floor by either one of the users. This sound will be referred to as the “touch down sound” from here on. This sound icon plays a drumbeat to indicate that a collision occurred. The last interaction sound was also used for the dropping event, but was instead heard when an object was dropped on top of another one. This sound will be referred to as the “collision sound” from here on. This sound icon also plays a drum-sound, but with a reverb applied to clearly differentiate it from the touch down sound. All of the above mentioned sounds were played in 3D to give a rough idea about the distance to the events and their horizontal positions. Technically, this means that the volume and panning of the sounds depend on the Euclidian distance from the center of the environment to the sound source (the place where a user is interacting with an object). Note that the sounds are in 3D in the virtual environment only. During the studies a pair of speakers was used to present the sounds to the users and hence the sounds were presented in stereo mode. However, an estimation of horizontal distance and depth could still be perceived. In the earlier study the visually impaired pupils had no means of knowing what was going on unless haptic
guiding was used or the sighted pupils explained actions taken. This constrained the collaboration and eventually excluded the visually impaired pupil from at least parts of the work process. The hypothesis for the present study was that it would be easier for a visually impaired user to take part in the task solving process if he/she got constant feedback on actions taken.

The fourth sound was an earcon, since it bore no resemblance with any event in the real world (see (McGookin and Brewster, 2004)), representing the position of the sighted user in the virtual environment. This sound played three consecutive notes of the same kind and is called the "contact sound". This sound was also played in 3D, since the purpose of it was to make a visually impaired user aware of where in the interface the sighted user currently was. This sound was heard related to the visually impaired user’s proxy position. The sound was played when either user pushed the button on the haptic device at a time where he/she was not in contact with an object (then, the grasp sound is heard instead). Although this earcon was specifically designed to indicate the sighted user’s position, the other sounds described above also indicate approximately where a user interacts with an object.

Apart from the addition of sounds some other modifications were also made to make interaction in the interface easier. First, the “wooden” box in the upper left corner was removed to make sure that the objects in it were not confused with objects used for the assignments. Second, the objects’ positions were locked (infinite mass) as long as they were not grasped by a user. This modification made it easier for the blindfolded user to explore the interface without moving objects unintentionally.

Table 3 below summarizes all available functions in the modified dynamic application tested in this experiment.

Table 3. All functions available in the application tested in the experiment.

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>All parts of the interface can be seen by the sighted user from a fixed angle</td>
</tr>
<tr>
<td>Haptic</td>
<td><strong>Grasping</strong> – objects can be picked up when device button is pushed</td>
</tr>
<tr>
<td></td>
<td><strong>Dropping</strong> – objects are dropped when device button is released</td>
</tr>
<tr>
<td></td>
<td><strong>Guiding</strong> – sighted user can guide the visually impaired by either holding on to the same object or grasping the peer’s proxy.</td>
</tr>
<tr>
<td>Audio</td>
<td><strong>Grasp sound</strong> – sound heard when an object is picked up</td>
</tr>
<tr>
<td></td>
<td><strong>Touch down sound</strong> – sound heard when objects is dropped on the floor</td>
</tr>
<tr>
<td></td>
<td><strong>Collision sound</strong> – sound heard when an object lands on another object</td>
</tr>
<tr>
<td></td>
<td><strong>Contact sound</strong> – sound heard when a user pushes the device button, which indicates the position of the sighted user’s proxy</td>
</tr>
</tbody>
</table>

5.2 The experiment

As already explained this study was a direct response to the results gained from the study on collaboration between visually impaired and sighted pupils. In particular, this follow-up study aimed at exploring how additional audio cues would influence the collaboration. Following up on the results on task performance from the earlier study (comparison
between different types of guiding) this new study also investigated how task performance was affected by adding sound to the dynamic application described earlier. The following research hypothesis served as a basis for the study, regarding effects on task performance:

\[(H1) \text{Adding audio feedback into a collaborative visual/haptic environment will make task performance faster.}\]

Two more general research questions, regarding more qualitative measures (described further in section 5.3.2) were also derived:

- \textit{How can multiple sensory cues with haptic and auditory feedback improve mutual understanding in order to enrich communication and enhance coordination?}
- \textit{How can such sensory cues make joint task performance more efficient in a situation where one of the persons cannot see the workspace?}

Since task performance was one of the measures of interest the decision was made to conduct a formal experiment in which groups of users solved tasks together. The experiment used a between groups design to make sure that potential learning effects would not affect the results. Thus, half of the groups recruited used the modified dynamic application without audio cues and the other half the modified dynamic application with audio cues added. The choice to use the between groups design was the main reason why it was not possible to use visually impaired users as subjects in the study – more subjects were needed than could be found, within a reasonable amount of time, in the target group of visually impaired users.

### 5.2.1 Subjects

For this study 32 participants were recruited among the students at the Royal Institute of Technology. These participants formed 16 pairs – one sighted and one blindfolded in each. The person to be blindfolded was randomly selected. The participants were recruited in pairs to make sure that the participants in all experiment sessions knew each other. This was mostly to make sure that the participants would be comfortable in the unnatural setting and that they could collaborate and communicate effortlessly despite the fact that one of them was wearing a blindfold.

Since a large sample was needed for this between groups experiment, blindfolded sighted subjects were used as substitutes for visually impaired ones in this study. This has its limitations, especially with regard to the issue of plasticity discussed earlier – it has been shown that visually impaired persons have more developed senses of touch and hearing (Shimojo and Shams, 2001; Collignon et al., 2009) than sighted persons. Röder et al. (2004) also showed that persons born with a severe visual impairment react faster to haptic stimuli than persons who became visually impaired later in life. It is not at all trivial to judge how these issues would result in differences in task performance between visually impaired and blindfolded persons. An important fact to consider here, however,
is that the earlier study (Moll and Sallnäs, 2013) discussed in chapter 4 was performed with visually impaired subjects and almost the same haptic interface. Thus, results from that study can serve as a good baseline when judging results from the study involving blindfolded subjects presented in this chapter. The only things that differed between the interfaces used in the two studies, as described in section 5.1 above, were that the blocks were not sliding anymore, that the wooden box was removed and that sound cues had been added. Even though the effects of adding audio have not been validated with visually impaired subjects, the influences that audio has on collaboration and communication can still be assessed. It is also important to consider that most sounds used are interaction sounds confirming that events have taken place. I argue that the perception of the kinds of sounds used in the experiment would not differ much between visually impaired and blindfolded subjects.

5.2.2 Apparatus

The equipment used in this experimental study was almost the same as the one used in the study on collaboration between visually impaired and sighted students (chapter 4), since the application and focus of the research were similar. Again, a DELL Precision PC with two dual core processors and ReachIn API 4.1 and Visual Studio .NET 2003 were used. Mouse and keyboard were again used by the researcher for loading tasks. Even in this case two different haptic devices (PHANTOM Desktop and PHANTOM Omni) were used, for the same reasons as given earlier. The only piece of hardware that differed from the earlier setup was a pair of speakers used to present the added audio feedback. Figure 17 below illustrates the hardware setup schematically.

![Figure 17. The test setup used in the experiment. (Huang et al., 2012).](image-url)
5.2.3 Setting

A screenshot from the video recording of one of the test sessions is shown in figure 18 below.

![A screen shot from one of the test sessions.](image)

As can be seen in the figure the participants sat beside each other at a table and once again the person who could not see was the one using the Desktop with better resolution. (See discussion about the difference between the haptic devices in section 4.4.2.) As in the previous study the sighted person saw the graphical interface on the computer screen, but in this case the screen was standing. This arrangement proved to influence the means of establishing common ground, as discussed further in the result section, since the visual interface was rotated in relation to the haptic one.

The loudspeakers were placed symmetrically on each side of the screen and the two participants respectively. This arrangement, also illustrated in figure 18, surely affected the perception of 3D sound to some extent. However, as the results show, a rough judgment of horizontal direction and position could be achieved without problem.

5.2.4 Procedure and tasks

The experiments consisted of five parts, where the participants were active; (1) demo session, (2) training session, (3) test task 1, (4) test task 2 and (5) interview. These steps were always carried out in the above mentioned order.

After the usual introductory information provided by the researcher, a demo application was uploaded so that the participant to be blindfolded should get an opportunity to interact in an interface which he/she could see. By providing this opportunity it was made sure that both users could handle the haptic equipment and get an understanding of how to haptically navigate in 3D before the real test tasks were presented. This is important, since the experiment is measuring the effects of adding audio. The training interface was a demo application presenting boxes with different haptic effects applied and thus it had nothing to do with the actual test application. None of the functions for lifting/dropping objects or guiding were available during the training.
After this introductory demo session, one of the participants was blindfolded after which a training task was uploaded in the dynamic application in which the real test tasks were going to be solved. The training task was the same as the one illustrated in figure 9 (page 55) and used in the study involving visually impaired and sighted pupils. In the training task the participants practiced the handling of objects and the guiding procedures described earlier. Half of the groups, who were going to solve the test tasks with sound feedback provided, were also introduced to the different sound cues. These groups also practiced on identifying the different interaction sounds and on using the localization feature.

After it had been made sure that both participants could use all functions effortlessly and as intended, the first test task was uploaded (figure 19).

![Figure 19. The first test task, as it looked directly after start-up. The users are represented by a blue and red sphere, respectively.](image)

In this task the participants should build a table with the building blocks provided;

In this assignment eight cubes of size 2x2 cm² 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 12x6 cm². 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 at the respective corners of the table so as to give a good-looking impression. 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 touch your own cubes.

The rationale behind this first assignment was to start with a simple idea and create a situation which made it necessary for the two participants to work together. Both participants read the written task description before one of them was blindfolded again and the task was loaded.

When the participants were in consensus about the task being solved they were presented with the second test task (figure 20);

On the floor there are now seven building blocks of different sizes. Three of these are cubes with volume 8 cm³ and the other types of objects have volumes of 8 cm³ or 12 cm³.
respectively. None of the objects can be turned. Your assignment is to build a cube with volume 64 cm$^3$ by using all of the building blocks.

![Image of building blocks](image)

**Figure 20.** The second test task, directly after start-up.

Through this test task it was possible to investigate the means of establishing common ground about the different building blocks, their positions and the orientations, in order to be able to reason about how to solve the rather complex task.

When both participants were in consensus about the task being solved they were then interviewed together. All pairs of participants were asked questions about their perceived level of common ground and awareness and about general usability. The pairs which solved the tasks in the visual/haptic condition were also asked if they thought sound feedback would be an aid and which type of sound cues they thought they would benefit from.

### 5.3 Analyses

The qualitative analyses were, as in the study with pupils described in chapter 4, based on video recordings and interviews. Both these data sources were analyzed using a content analyses approach, resulting in emergent themes of special interest. The general procedure followed for analyzing the recordings was the same as in the first study, and is explained in section 3.1 about analyzing dialogues. The categories relevant for this thesis are listed in the next two sub-sections.

The interviews were also transcribed in their entirety. After transcription, the different parts of the interviews were summarized and sorted under the respective questions asked. This was done for each group. Answers from the haptic and audio groups were compared in a last iteration.

In this experiment both quantitative and qualitative data were gathered and, as opposed to the earlier study with visually impaired children, a hypothesis was formed prior to the study. Note that two articles have been published with results from this study; one mainly presenting the quantitative results (Huang et al., 2012) and one the qualitative results (Moll et al., 2010). In the remainder of this section the data gathered from and the measures used in this study will be described.
5.3.1 Quantitative measures

The quantitative measures used will be described here. Note that “Time to perform task” is the only one of the listed measures which was subject to statistical analysis. All other quantitative measures were found to be important during the iteration of the observation material and were thus not planned at the outset of the study.

- **Time to perform task**
  The time it took for the respective task to be solved, measured in seconds. The timer started counting when the task was loaded and stopped when both users agreed that the result was satisfactory. This measure was subject to statistical analysis and corresponds to hypothesis H1, and sheds light on whether the added sound cues have an effect on task performance.

- **Number of verbal guiding operations**
  The total number of verbal guiding operations in the haptic and audio groups respectively. The definition of verbal guiding is the same as the one used in the first thesis study. This measure was not subject to statistical analysis and is in that sense indicative. It sheds light on how the added audio feedback influences the verbal communication between the participants.

- **Number of clarifying questions**
  The total number of clarifying questions asked by the blindfolded and sighted participant in the haptic and audio groups respectively. A clarifying question is in this case defined as a question about actions currently taken by oneself or the other user (e.g. “What are you doing now” or “Did I drop it on the right spot”). These values are also indicative and shed light on how good the sound cues are at conveying appropriate awareness information, since these kinds of questions reflect uncertainty.

- **Number of successful/unsuccessful uses of the contact sound**
  The total number of times the contact sound was successfully and unsuccessfully used, respectively. The contact sound is successfully used if the blindfolded participant can navigate to the place where the sighted person’s proxy is located with only minor additional need for verbal directions for fine-tuning the position. This measure is also indicative and sheds light on the effectiveness of the 3D contact sound.

- **Number of successful uses of the drop sounds**
  The total number of times where it was evident, from the dialogue, that either one of the users made use of the awareness information provided from the touch down sound and the collision sound. A successful use of a drop sound occurs when it is obvious from the dialogue or the behavior that the blindfolded or sighted participant can use the sound to understand if an object is dropped on the floor or on another object. This measure is also indicative (no statistical analysis) and sheds light on whether the sound cues were really used.
5.3.2 Qualitative measures

Here, the qualitative measures considered in the study will be presented. In the analysis the categories “Usability” and “Modality” were used and these are also presented in the resulting qualitative article (Moll et al., 2010). These categories are too general for this thesis since they relate to all other categories used. This is why those categories will not be described here.

- **Common ground**
  Common ground was defined in the same way as in the previous study; “a state of mutual understanding among conversational participants about the topic at hand” (Clark and Brennan, 1991). In this case the focus lied on different means of establishing common ground in the haptic and audio groups respectively. How these differences are shown in the dialogue between participants is of utmost importance in this study.

- **Awareness**
  Awareness was defined in the same way as in the previous study; “an understanding of the activities of others, which provides a context for your own activity” (Dourish and Bellotti, 1992). In this experimental study the focus was on the additional awareness that the audio cues provide. The most important issue considered in this study was how access to audio cues influenced the collaboration and most of all the communication between the two participants.

- **Guiding**
  The means of and reasons for providing verbal, haptic and audio guiding to the blindfolded participant. Verbal and haptic guiding were defined in the same way as in the previous study. Audio guiding was for this study defined as “a conscious use of the contact sound to define a point to which the blindfolded participant should come to”. Guiding is the most important qualitative measure considered in this study since it sheds light on how different modalities can be used for communication.

- **Social presence**
  If, and in that case how, the participants felt that they were together in the same environment solving a shared task. How haptic and audio, respectively, supported this feeling is of special interest here. Note that the definition of this category is wider than the one used in e.g. telepresence theory, which mostly relates to the affordances of communication media used.

5.4 Main results

In this section both the quantitative and the qualitative results will be provided and discussed. As in section 4.6, the qualitative results will be illustrated further by dialogue excerpts. In these excerpts the sighted participant is denoted S and the blindfolded B.
More results, with more peripheral relevance to the focus in this thesis, can also be found in the resulting articles (Moll et al., 2010; Huang et al., 2012).

### 5.4.1 Quantitative results

A one-way ANOVA was performed to investigate if there was a significant difference in task completion time for the two tasks between those who used and did not use the audio feedback respectively. The analysis showed that there was a significant difference for task 1, as seen in the table below:

<table>
<thead>
<tr>
<th>Table 4. Quantitative results on task performance, measured in seconds.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haptic feedback</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td><em><em>Task 1 (n=14, F=5.0</em>)</em>*</td>
</tr>
<tr>
<td><strong>Task 2 (n=14, F=0.006 (p=0.94))</strong></td>
</tr>
</tbody>
</table>

* = significant at 95% level

As can be seen from the results the hypothesis was confirmed for the first task at the 5% significance level and hence it was shown that task performance was faster when the sound feedback was available.

As seen in table 4 above significance was only reached for the first task. This is probably because this task was simpler to understand – the challenge in task 1 lied on actually carrying out the task and not on figuring out how to do it cognitively. In the second task, the participants should solve a rather challenging puzzle and thus the task completion times were more affected by the task difficulty than by the availability of sound cues.

The results from the other quantitative measures, which were not statistically analyzed, are presented in table 5 below.

<table>
<thead>
<tr>
<th>Table 5. The rest of the quantitative results, not subject to statistical analysis (the unit is numbers).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure Group</td>
</tr>
<tr>
<td>Haptic</td>
</tr>
<tr>
<td>Audio</td>
</tr>
</tbody>
</table>

The results in table 5, although not tested for significance, strongly indicate that the access to sound cues make a positive difference in a number of respects. The different measures are explained in section 5.3.1 about the quantitative measures. Many of these values have a direct bearing on the communication between the participants. First of all, verbal guiding is used a lot more by the participants in the haptic only groups. This indicates that the access to additional sound cues decreases the need for verbal guiding.
and this in turn has implications for the collaboration, which will be described more thoroughly in section 5.4.2 about the qualitative results.

Even when it comes to clarifying questions the results are clear. These types of questions, reflecting some kind of uncertainty about actions taken by oneself or the other user, occurred a lot more often in the haptic only groups as can be seen in table 5. Questions about the actual interaction and the course of events take unnecessary time from the actual task solving process and in this respect these results are important. A decrease in clarifying questions also shows that both participants are more confident with their own and the other’s interaction with the environment when audio feedback is included, despite the fact that the cues used were rather simple.

The last quantitative results tell more about the effectiveness of the sound cues. As can be seen in table 5 the success rate of using the contact sound was high, meaning that it could indeed be used to guide the blindfolded user to a specific place. The contact sound was used as a means for the sighted person to provide guidance to the blindfolded person in all these cases. According to the results from the interviews, the blindfolded participant never used the sound his-/herself to find out where the sighted person was. The drop sounds were also used successfully in several occasions.

5.4.2 Qualitative results

In this section the most important results from the study will be provided and elaborated on. For more results and examples of more peripheral relevance to the thesis focus, the reader should consult the article (Moll et al., 2010).

5.4.2.1 Providing guiding

Means of providing guiding to a visually impaired peer played an important role also in this study, since it is one of the main themes in this thesis. Three different types of guiding could be provided by means of the functions accessible to both users – verbal, haptic and audio. To illustrate the main points three examples will be provided below on the occurrence of each type of guiding.

Example 5 (verbal guiding):

S:  Ok, right,... right, right,... left,... feel,... you can feel the corner
[B moves to the upper right corner]
S:  ... and then go to the left a bit... stop!... down... left, left,... down,... go down till you feel the ground... and then go forward until you feel the wall...
[B moves the block to the upper wall, sliding it on the floor]
S:  ... and right until you feel the boxes,... so, yeah,... let go

Example 6 (haptic guiding):

[S grabs B’s proxy]
[He drags B to the beginning of the L-shape]
S:  Now, here we have an L-shape...
[S drags B to the top of the shape]
S:  
... this is the top.
[S now drags B back and forth on the L-shape's north-southern part a few times]
[He then drags B to the east, until the shape ends]
S:  
Ok, and this is the bottom right... and then we have this cube that is taller than the others
[He drags B up and down on a tall block placed beside the L]
S:  
We have another one just like it

Example 7 (audio guiding):

S:  
Pick up a new cube
[B locates a cube on her own]
B:  
That one?
S:  
Yeah...And then you can move here...
[S uses sound to show the way]
[B navigates to a place slightly above the intended one]
S:  
Ok, down a bit..., down..., stop
[B releases]
S:  
You can try to pick up the cube that's here...
[S uses the contact sound again]
[B navigates to the exact place in a few seconds]

When comparing the examples 5-7 above the potential of haptic and audio guiding is made clear. Example 5, in which the sighted user verbally guides the blindfolded one on where to place a block he is holding, illustrates the core problem of verbal guiding especially in situations in which one of the users cannot see the environment. Guiding a person, who is moving around in 3D, by verbal directions, is obviously cumbersome and it takes time and energy from the actual task of composing a large cube. This example of verbal guiding is typical for this study and several similar examples could also be found in the study with visually impaired pupils discussed in chapter 4.

Example 6, illustrating haptic guiding by means of grabbing the other’s proxy directly, shows the potential of utilizing a haptic guiding function. Instead of verbally directing the blindfolded user around the blocks described, the sighted user “grabs” his peer and quite physically shows the blocks as he is describing them. This approach saves time and is utilized in a straightforward way. The approach followed here also provides an interesting example of an attempt to achieve a common ground about the different building blocks from which a cube should eventually be composed. The possibility to guide haptically, as illustrated in example 6 enables the use of deictic referencing like “here” and “this”, as the blocks are browsed through. Example 6 shows how the function of grabbing the other’s proxy can be used in the collaboration, but even though no direct comparison can be made between visually impaired and blindfolded subjects, the examples provided of holding on to the same object in chapter 4 are also typical for the study presented here.
Last, example 7 illustrates a typical use of the contact sound, the main purpose of which is to give the blindfolded user an impression of where the sighted peer is located in the shared environment. In the example, the contact sound is used two times to guide the blindfolded user to the sighted user’s location, where blocks to be picked up are located. As in example 6, it is clear that the contact sound, as it is used in the example, decreases the need for providing verbal guiding and in this way makes the collaboration more efficient and straightforward. Again, the use of deictic terms is made possible through the use of sound feedback. In most cases the sound could be used to direct the blindfolded to the correct place in the horizontal direction quickly, but as the example shows there could be an additional need for verbal guiding in the forward/backward direction.

The examples provided above show that haptic and audio feedback can play an important role in the collaboration between a sighted and non-sighted user. It was already clear from the study with visually impaired pupils that the haptic guiding functions were important for the efficiency of the collaboration and the results in the study presented here confirm those findings. The results also show that the added audio function used for guiding purposes serves a similar role, even though there were some deficiencies in the design of the contact sound. The design of audio cues is discussed further in section 5.5.2.

5.4.2.2 Influences on verbal communication
In this study, as well as in the study presented in chapter 4, the modalities’ influences on verbal communication is one of the most important themes considered. In example 6 and 7 above some of these influences are illustrated – both haptic and audio guiding functions enable straightforward means of providing deictic references. By using the guiding functions provided, the focus was more devoted to the actual tasks than the constant need to give verbal directions about navigation. In this way the guiding functions not only influenced the dialogue but also eased the collaboration considerably and increased the efficiency of it.

As in the study on collaboration between sighted and visually impaired pupils (chapter 4) the means of establishing common ground through haptic and audio were also studied in detail here. Example 8 shows a typical example of how haptic feedback can be used while forming a common ground about the building blocks needed to build the composed cube in the second test task. The example also illustrates pointing in the interface. On several occasions the blindfolded participants pointed out different objects by repeatedly hitting the object from above with the proxy.

Example 8:

S:  
I'm just trying to find out which blocks are the highest...
[S feels around and ends up on a high block]

S:  
...yes, the blue ones are highest

B:  
Ok

S:  
and the other one...
[B moves around and ends up on a long block]

S:  
there are two... you're at a red one now...

B:  
That one?
[He points repeatedly to the long block to clarify]
[He then moves along the long and short edges, respectively]

**B:** quite short... quite long

**S:** Yes, and the height...

[B moves up on the box and down again a few times]

**B:** short!

**S:** Yes, the low ones... The other one is... no, up in the corner

[B navigates to the forward right corner]

**S:** There it is

**B:** Ok

**S:** This is the other way...

**B:** Yes, turned 90 degrees left,... ok

The above example clearly illustrates how the haptic sensation is used in the grounding process. In large, the sighted user focuses his explanations on what the blindfolded one currently feels. In this way the haptic feedback provides a shared focus even though both users are not using the haptic equipment for the time being. It is also apparent from the above dialogue excerpt that the users most of the time are discussing the different building blocks as if they both could see the entire scene and this is again due to the shared focus provided by the haptic feedback. Although not related to haptic or audio feedback, it is worth noting that the participants in the above group partly used colors to refer to the objects. This does not seem appropriate given that one of the users could not see, but later on during the task solving process both users referred to the objects by color in a straightforward way. In most groups, however, the participants referred to the objects by means of their dimensions. The following excerpt, from the interview with the group from the example above, illustrates why it worked to use color for object referents (again, illustrating the use of haptic feedback):

**S:** Yeah, I had to use some way to,... instead of using the dimensions of the boxes which was quite hard to explain I used colors just to assign a color with a dimension.

**B:** Yeah, I remember you told me "this is the red one" and I could feel it first and I remember the shape.

Since object dimensions or positions were most important during the early stages of grounding, haptic feedback was most utilized in this phase, by both the haptic-only and the audio groups. Sound was, in general, not used much during grounding. This does not come as a surprise since the most important function of the sound feedback was to provide awareness of actions taken in the interface.

### 5.4.2.3 The significance of audio feedback – about providing confirmation

When it came to establishing a common ground about the task and the objects at hand, as described above, there were no apparent differences between the haptic-only and audio groups in the experiment. This was much due to the fact that haptic feedback was more suitable for grounding object orientations and dimensions, as discussed earlier. The true benefit and potential of the audio feedback came to light when studying awareness aspects. The possibility of the sound feedback to convey awareness information about the
actions performed by oneself or the other user explains much of the difference between the haptic-only and audio groups.

Example 9 below, in which the blindfolded user tries to place one cube on top of another to compose one of the table legs in the first test task, illustrates one of the most important points here.

Example 9:

[B puts the cube down, floor collision heard]
[B puts it down again, floor collision again]
[B lifts and places the cube on another one, object collision heard]

B: Does it look nice?
S: No...
B: But I know it’s on top...

In the above example the blindfolded participant is trying to do the task of placing one cube on top of another by himself. After two failed attempts, in which he places the currently held cube on the floor, he manages to place the cube on top of another one that he placed by the forward-facing wall about a minute earlier. He uses the difference between the two sound icons representing collision with floor and collision with another object, respectively, to come to the conclusion that the subtask of building a table leg was soon to be completed. That the blindfolded user attended to the sound is evident from the example, since he poses the question “Does it look nice?” after he hears the correct sound. This is not the most important finding from this example, however. The example also clearly illustrates that the sound feedback made it possible for the blindfolded participant to do a task on his own. Thus, the sighted and blindfolded participants can work partly in parallel and thereby more efficiently. This is made possible since the sound gives feedback to the blindfolded participant about his own actions.

The other key finding, related to sound and awareness, is illustrated in example 10 below.

Example 10:

[B is exploring one of the long blocks by touch]
S: Ok, you can pick that one up again
[B continues exploring the object for about 10 seconds]
[In the meantime S moves another block and tries to fit it into the solution]
[B grasps his long block and the grasp sound is heard]
S: Ok
[S starts guiding to the correct place]

The above example illustrates that even the sighted participants can make good use of the sound feedback generated from the blindfolded’s actions – S starts to guide the blindfolded only when he knows that the block has been picked up. The awareness information given by the grasp sound in this case makes it possible for the sighted to
work on a parallel task until the time has come to guide the blindfolded – the sighted user
does not have to wait and look at what the blindfolded is doing.

These findings from the audio groups are interesting to compare with findings from
the haptic-only groups. In the haptic-only groups the blindfolded participants hardly ever
tried to solve subtasks completely on their own. The reason for this is most probably that
they lacked the awareness cues that the sound icons provided for the audio groups. In the
haptic-only groups the blindfolded participants did not get any information (unless
provided verbally) about changes made by the sighted person and they did not receive
any feedback on their own actions. The blindfolded participants’ situation in these groups
thus resembled the situation for the visually impaired pupils described in chapter 4.
Questions like the following occurred in almost all haptic-only groups (note that these are
eamples of the clarifying question being parts of the quantitative data gathered).

Did I drop it?
Did you pick it up?
What is going on?/Are you doing anything?

These types of questions, asked by both sighted and blindfolded participants, almost
never occurred in the audio groups, and are clear indications of lack of awareness.
Obviously, the haptic feedback did not give enough confirmation about own actions
taken and this often resulted in the blindfolded users in these groups being passive,
waiting for instructions from their sighted peer. This explains the quantitative results
provided earlier, in section 5.4.1, to a great extent. The lack of awareness of one’s own
actions in the haptic-only groups forced the blindfolded participants to depend on the
sighted peers and parallel work was thus not an option. This fact, in combination with the
restriction that the participants could only touch their own set of cubes in the first test
task, most probably resulted in the longer task completion times. The awareness issues
discussed here also explains why clarifying questions were asked a lot more often in the
haptic-only groups than in the audio groups.

Although most comments in the interviews regarding the sounds’ roles in conveying
awareness information were positive, the sound model implemented was not
unproblematic. As stated earlier the contact sound was implemented as a 3D sound, but it
proved difficult to locate the sound source in the forward/backward direction. Some
groups found this annoying, since they could not use the sound entirely as intended.
Some blindfolded participants were also irritated by all sounds heard when the sighted
peer interacted with the objects. This was much due to the fact that there were no details
provided about which objects were moved. In cases when the blindfolded did a subtask
on his/her own the sounds generated from the sighted’s interaction also did not matter
much and were more of an annoyance. These comments aside the results show that the
idea to implement sounds was a success, with regard to the overall research questions and
hypothesis, and the following comment formulates the core benefit of the sounds in a
clear way.

With no feedback it wouldn't be possible but without sound I think it would be
possible but it wouldn't be as easy, and it would be kind of boring for me because
I'd have no idea of what he was doing if I hadn't had the sounds.
5.5 Lessons learned

Since this is a follow-up study on the evaluation presented in chapter 4 most of the lessons learned from the study presented here relate to the difference between the results between the two studies. The single most important point to be made here is that the hypothesis stated at the outset of the project was confirmed – adding audio to the visual and haptic application made task performance faster. The significance of this result for this thesis and for research in general will be elaborated on further below.

5.5.1 Deepened insight into haptic communication

The results of the study presented in chapter 4 gave some valuable insights into how haptic feedback can be used for communication between collaborating participants and how that kind of feedback can influence the dialogue. One of the main benefits of haptic feedback identified during that study was the ability to provide guiding functions. These functions, in turn, proved to influence the dialogue in that fewer words were spoken during the haptic guiding operations. The potential of the haptic feedback when it came to establishing a shared frame of reference was also made clear (Moll and Sallnäs, 2013).

Although all of these results are important they originated from a rather small study and this is the main reason why a more extensive study was performed with a larger sample. The follow-up study generated important results in a number of respects and this section will focus on the results related to haptic communication. Despite the fact that there are clear differences between visually impaired and blindfolded subjects (see discussion in section 5.2.1) most of the results regarding haptic communication were confirmed in the follow-up study.

Once again, the haptic guiding functions proved to play an important role in the collaboration. Example 6 from section 5.4.2.1 presents one of the core benefits of using haptic feedback for guiding and several similar examples were found as well. It is clear that the implemented guiding functions eased the collaboration considerably both when it came to establishing common ground about the objects at hand (example 6) and when it came to helping the blindfolded participant to solve subtasks more efficiently. Example 6 and similar examples also illustrate the idea that the guiding functions can be used to convey deictic information haptically. While the sighted participant drags the blindfolded’s proxy around the different building blocks he describes the different parts of the blocks felt. At the same time the blindfolded participant gets information about directions and lengths through the haptic feedback – information which does not have to be stated explicitly. Thus, haptic feedback is used to provide meaningful and important information in parallel to the verbal descriptions. A lot of this information is deictic in nature.

It is also clear from the study results that the focus in the dialogues between the participants is influenced when guiding functions are used and this is in line with earlier research as well (Moll and Sallnäs, 2013; Bivall-Persson et al., 2007; Chellali et al., 2011). In this case it is clear that the haptic feedback influences the focus in that the participants can concentrate on the actual task to be performed and not on directing the
blindfolded participant. As already concluded this finding confirms the results by Moll and Sallnäs (2013) and is one of the most central findings in the thesis.

By this follow-up study the knowledge about how the haptic feedback plays a role in deictic referencing was also deepened. Example 11 below should make the points clear.

Example 11:

S: *Everything that is 3 high has to be combined with 1*
B: Yeah
S: *And there are two of them right now*
[B moves up to the front right corner]
B: *One in the corner here*
S: Yes
B: *And the other one should be here*
[B moves around on the floor in the backward left corner of the soon to be cube]
S: No
B: *Closer to me... so that this one can be on top...*
[B points to a long block more than 1 dm below!]
[B then moves with his proxy back and forth on the left side of the soon-to-be cube to show what he meant with “on top”]
S: *But we have only two of those that are 3 high*
B: Yeah
S: *...but we have two of the flat ones*
B: Yes... I think I have a pretty good picture of how it should be
S: Ok, so where do you want to put it? There is something wrong now, one is standing on the floor and the other one on the long one
[B points to the one standing on the ground]
B: I want to put it closer to me... It should be close to me and next to the lying one in the corner
S: Next to the lying one?... What you currently have is one flat one lying in the corner
B: Yes, and on top of that there should be one three high and one two high
[That is the current situation]
S: Yes
B: And then, going towards me from the 2 high you should have a 3 high standing on the ground
[Now, the sighted sees the solution and the remaining blocks are placed]

Example 11 is one of the best examples of how haptic feedback can contribute to a discussion, among the two studies presented thus far. Similar, but shorter, discussions like these occurred numerous times in the transcripts. It is clear that the focus of the discussion, without any exception, is on the task to be solved and especially on how the different blocks, given their dimensions, should fit together to form the resulting cube. No verbal guiding is needed. It is also clear that both participants are taking active part in the discussions and that the blindfolded one is the one taking the initiative and who
actually solves the problem. The haptic feedback is used to establish a common frame of reference and a shared focus in several instances in the above example. Words like “here” and “there” are used fluently and accurately and the blindfolded participant often pairs these references up with pointing repeatedly at or sliding against the object. In the example the blindfolded participants use haptic feedback in combination with verbal descriptions to illustrate a strategy, but in several other examples the haptic feedback provided a shared focus of discussion in that the sighted participant talked about what the blindfolded currently felt.

After this study there was no doubt that haptic feedback can influence the collaboration and especially the communication between sighted and non-sighted participants in several respects, while working with the PHANTOM haptic feedback equipment. This insight resulted in a publication formalizing the ideas regarding haptic communication presented this far (Moll and Sallnäs, 2009).

5.5.2 The difficulty and necessity of providing accurate feedback

Up until now most discussions in this thesis have concerned collaboration and communication, but this does not mean that the technical aspects of the work performed is unimportant. There are several important points that can be made after the two studies presented thus far.

It was clear from the study with visually impaired and sighted pupils that lack of awareness in a dynamic environment was detrimental, regardless of the fact that the implemented functions were usable and served their purposes. The only feedback on actions taken in the environment was the haptic sensation that something had been moved. This lack of awareness information put the visually impaired pupil in a passive state and he/she had to rely a lot on the sighted peer. The fact that the objects were sliding on the floor (a conscious decision to make it easier to fit blocks together) made it even harder for the visually impaired to scan the environment. As it turned out this design choice affected the collaboration and then especially the inclusion of the visually impaired pupil in a negative way. The focus of the study with pupils was to evaluate the potential of haptic feedback, in particular, on inclusion of the visually impaired pupil. From the results it is obvious that the haptic feedback was not enough for conveying the necessary information needed to solve tasks on equal ground and from the discussion in section 4.7.2 it is quite clear that haptic feedback, in this case, is not suitable for conveying the missing awareness information.

The issues raised here are interesting to compare with results from the haptic-only groups in the follow-up study with sighted and blindfolded students. As discussed in the result section a lot of clarifying questions were asked about the actions taken by both the sighted and blindfolded participants and these types of questions are strong indications of lack of awareness. Thus, many of the problems identified during the first study were confirmed by the second study, by the groups who did not get audio feedback. Even though this was expected, the results are now confirmed by a larger sample. Since the blocks were locked in position as long as they were not grasped, in the follow-up study, the sighted and blindfolded participants had an easier job reaching a common ground about the orientation and dimensions of the different blocks used. In the study with pupils
in elementary school (chapter 4) the blocks moved when the visually impaired pupil should explore them, unless the blocks were placed against a wall or in a corner.

Results from the audio groups show that the implemented sound feedback made all the difference as can easily be seen from the examples provided in the result section 5.4. While the original implementation, used by half of the groups in the follow-up study, put the blindfolded participant in a passive state in several cases and caused verbal communication not directly relevant for the task at hand, the blindfolded participants in the audio groups often solved subtasks on their own and hardly ever asked questions reflecting a lack of awareness. In this respect it is safe to conclude that the rather simple sound model served its purpose. The comparative study performed shows that a few simple sound icons can make a big qualitative difference for the collaboration between a person who can and a person who cannot see.

Even though the results gained were positive there were still issues that influenced the collaboration and the communication between the participants in subtle ways. The biggest problem, brought up to discussion earlier, was that it was hard for many of the blindfolded participants in the audio groups to understand exactly what was going on in the interface. They could hear approximately where an event occurred but they did not know which of the available objects the sighted peer interacted with. As a result some of the blindfolded participants reported that they filtered away most of the sound feedback that came from the sighted person’s interaction. Note that this does not mean that the sound system was a failure – almost all blindfolded participants reported that they felt comfortable with knowing that work was in progress and they could really make use of the feedback they got from their own actions. The sounds implemented had a positive effect both when it came to quantitative and qualitative aspects. The result that most feedback was filtered away can also be related to the results gained by Baharin and Mühlberger (2010) who concluded that we actually need to listen to a sound to be able to deduce a meaning from it. The results from the thesis study indicate that the participants were able to use the sounds for gaining awareness information if they really focused on them. Our ability to focus when we really need to and filter away sounds when we do not need them is important in this kind of interface in which each and every action produces a sound.

Except for the lack of detailed information in the audio feedback there was also an issue with precision brought up by many participants as a negative aspect. Judging from where a particular sound came from (where a particular action was taken) was fairly easy in the left-right direction probably because stereo speakers were used to play the sounds. When it came to the forward/backward direction it was a lot harder and this was reflected by the fact that audio guiding through the contact sound often required some extra verbal guiding to fine-tune the blindfolded’s position. This deficiency resulted in negative comments, but the influences on the overall collaboration and communication seem to be minor.

All in all, the results show that the changes made to the dynamic application between the two studies were effective and they also shed light on the importance of presenting accurate feedback to the users. It has been clearly shown that the collaboration between a sighted person and a person who cannot see is influenced when additional awareness information is added. One of the biggest impacts is that the person who cannot see moves from a more passive state to an active state and this affects the overall collaboration as
shown by many of the examples presented in this chapter. However, when judging the results it is important to have in mind that more would be needed to produce a more usable sound system. The issues of design and resulting design implications will be discussed more in section 7.3 about technical considerations.

### 5.5.3 The audio could make task performance faster

As confirmed by the quantitative analyses on task performance, and supported by the other quantitative data, task performance is faster when sound is added to the interface evaluated with visually impaired and sighted pupils (see chapter 4). As the qualitative analyses show, the most probable reason for this is that sound provides the earlier lacking awareness information – information that is absolutely necessary for efficient and successful collaboration to occur.

Not many earlier studies have been performed regarding the effects of audio on task performance in situations when sighted and non-sighted users solve tasks together in a haptic and visual interface. Lee and Spence (2008) showed that using trimodal input is significantly superior to using unimodal input, but did so for sighted users and not based on tasks involving joint manipulation of objects. Yu and Brewster (2003) also showed that task completion times were significantly faster when blind subjects should solve tasks based on a bar graph, when sound was added to convey extra information not easily conveyed through haptic feedback. However, the situation considered in the current study, in which two users collaborate on solving tasks using haptic and audio to communicate with each other, is quite different.

After the two studies discussed thus far in this thesis it was quite clear that both haptic and audio feedback influences the communication between a sighted and non-sighted user and that task completion times are positively affected by the addition of audio feedback. The next logical step was to investigate which of these facts hold true for collaboration between two sighted users. With the study of collaboration between sighted and blindfolded the thesis project also took a turn in the quantitative direction. This is one of the main reasons why the next study concerned collaboration between sighted users and included task completion measures for more modality combinations, while still keeping the focus on collaboration and communication.
6 Scanning and placing in an abstract multimodal interface

Up until now the thesis studies have only focused on collaboration between sighted and visually impaired or blindfolded users. Even though the potential of haptic and audio feedback for providing a shared interface is highlighted in these cases one could not help wondering if the same would hold true for collaborative task solving between sighted users. Additionally, my interest for how different combinations of modalities influence the collaboration became more and more pronounced over time. This is why this next experimental study focused even more on performance measures and comparisons and involved only sighted users. However, to keep some connection to earlier studies most of the target areas used in this study were visually occluded. While the earlier studies showed that haptic and audio functions can be utilized for communication, this study showed that the combination of modalities influences the collaboration and communication despite the fact that no functions for e.g. guiding were implemented.

This study was performed within the project MCE (Multimodal Collaborative Environments). As before, I developed the haptic part of the application to be tested. Sten-Olof Hellström and Fredrik Winberg collaborated in developing the sonification since I do not have any expertise in the area of sound design. The study was planned by the three of us and Eva-Lotta Sallnäs Pysander and it was carried out by me. I also performed most of the quantitative and qualitative analysis and was main responsible for the resulting article (Moll et al., 2013).
6.1 Scanning for and placing objects in multimodal interfaces

In the studies presented in chapters 4 and 5 a lot of focus has been placed on how haptic and audio functions can be used for communication and guiding in particular. These studies show that there is a potential for haptic feedback to be used for communication in multimodal collaborative environments, in which sighted and people who cannot see solve tasks together. There are many interfaces, however, in which sighted people are collaborating, where the target areas are visually occluded. This is the main reason why this new study was conducted – to see if access to haptic and audio feedback can influence collaboration and communication even in cases where both collaborators are sighted. The earlier studies also included specific haptic and audio functions used for communication. This new study did not include any such functions, which made it possible to investigate if the combination of modalities can influence collaboration and communication even though functions for communication are not present.

For this study an abstract interface based on sorting of information was developed. In short, the interface contained a bottom plate on which a topography of cubes were placed in a 10x10 grid of “cells”. Two users should collaborate in finding cells, on the plate, where there were currently no cubes and fill these cells out with cubes falling from above. Section 6.2 below will explain the interface and the functions available in much more depth. In interfaces containing visually occluded target areas, like the one used in this study, other modalities than vision need to be utilized for task completion to be possible. The two actions that need to be performed are scanning and placing, as defined below:

- **Scanning**
  The action of finding the target areas, or empty cells in this case, by means of the modalities currently supported.

- **Placing**
  The action of establishing the exact position of the empty cell found during scanning and the action of filling the cell out with a falling cube. Note that this definition should not be confused with Clark’s (2003) definition of “placing-for”, which was introduced in section 2.2.1.3.

On the general level, scanning through a complex scene haptically is difficult (Kahol and Panchanathan, 2008) especially if the exploration is done through a probe (Lederman and Klatzky, 2004). The main reason is that it is hard to get a general overview of the scene. On the other hand haptic feedback has shown to be an aid when judging distances and it can also enhance the 3D perception of a scene (Hale and Stanney, 2004). This indicates that haptic feedback is suitable when e.g. establishing exact positions of interface elements in a scene. The sense of hearing is the strongest sense when it comes to detecting temporal changes (Lederman and Klatzky, 2009) and this indicates that scanning of a complex scene by means of audio feedback could be efficient provided the audio feedback is appropriately designed. As can be deduced from this short theoretical discussion the senses of touch and hearing have different strengths providing advantages in either the scanning or placing phase. As a consequence it is natural to believe that the
combination of modalities influences task solving strategies, but it is not clear how (especially in a collaborative setting).

In the study presented in this chapter it was investigated how the combination of modalities influenced task solving strategies for scanning and placing during joint task solving in the interface described in section 6.2 below. As before the focus is on communication, and thus means of establishing common ground, negotiating task solving strategies and grounding results on actions performed.

6.2 Developing the interface

Since the objective of the study presented in this chapter was to see how haptic and/or audio feedback influenced the communication among sighted users the choice was made to develop a visually demanding interface with visually occluded target areas. This forces the users to depend on other senses than vision and the effects of haptic and audio feedback respectively will be easier to isolate.

The choice was made to develop a kind of game interface illustrated in figure 21 below.

![Figure 21. The application used in the third thesis study.](image)

The developed interface consisted of two main parts – one main program responsible for user input and visual and haptic feedback and one sound engine responsible for the implemented sound model. These two parts ran on the same computer and communicated via the UDP protocol. The application is meant to be used by two users in collaboration.

The main program’s graphical interface consists of a flat surface on which piles of cubes are placed (see figure 21). The red surface on which the cubes are placed is divided into 10x10 quadratic cells, each with sides equal to the sides of the cubes. Thus, all piles of cubes seen in figure 21 are positioned on their respective cells. When the program is
started 150 cubes are randomly placed on 90 of the 100 cells. Thus, when the user first sees the application running there is a complex topology of cubes placed all over the red flat surface. What the user does not see is that 10 cells are not covered by a cube. These cells will from here on be called empty cells and these serve as target areas during the experiment, as explained further on. There is never an empty cell on the row closest to the users. The angle from which the scene is experienced is locked and hence the users cannot change the viewpoint. The viewing angle can be seen in figure 21.

After the program has been started cubes starts falling down from above. They come one at a time starting above a random cell and fall down step by step until they either reach the flat surface or a pile of cubes. As soon as a cube has landed a new one starts falling down from a random location. Both users can control these falling cubes by means of the arrow keys on their respective keyboards. If one of the users presses the right arrow key, for example, the currently falling cube will move one cell to the right provided that the cube is not in the rightmost column. The whole idea with the application is that the users should make informed decisions about where to place these falling cubes in order to achieve a certain configuration of cubes on the flat surface.

Apart from keyboards users also interact with the application through the PHANTOM haptic devices used in the earlier studies discussed in this thesis. Through these devices it is possible to feel all interface elements – the red flat surface, the lying and falling cubes and the space between the lying cubes. There is also a clearly noticeable difference in friction between the cubes and the flat surface. By using haptic feedback to scan the environment the users are able to gather enough information in order to decide where to put a falling cube – they can find the 10 empty cells by scanning for positions with considerably lower friction and they can make sure that piles get a certain height by counting the spaces along the piles and filling up with falling cubes when needed, just to mention a few examples.

There are two types of sounds implemented in the sound model – shared and personal. Accordingly, there are sounds heard by both users, simultaneously, and sounds that are only heard by one of the users. Two of the shared sounds concern the currently falling cube and are meant to be an aid in determining the position of the cube. The first shared sound, here called fall sound, informs the user about how high above a pile the cube currently is. It is a constantly beeping tone which goes up in frequency as the cube approaches the pile (or the flat surface if there is no pile under the falling cube). By looking at the topology of cubes while listening to this sound, the user can tune in the position of the falling cube by e.g. moving it in relation to high piles. The second shared sound, here called touch down sound, is heard as a continuous tone when the falling cube has landed on top of a pile or on the red surface. The sound is heard for one second after which the next cube starts falling down from a random position. During the short period the touch down sound plays, the user has the chance to move the cube’s position if needed. The third and last shared sound is an ending sound. When all 10 holes had been filled out this constantly beeping sound was heard as a confirmation of task success.

The personal sounds are related to the users’ individual scanning of the scene. When a user positions the proxy over a pile of cubes a rapid series of \( n \) cliques is heard, where \( n \) represents the number of cubes in the particular pile. This sound will from here on be called the scan sound. When the proxy passes over an empty cell, where there is currently no pile of cubes, another scan sound is heard; the empty scan sound. This sound plays a
short scratching sound. By attending to these two sounds, the users can get a feeling for the topography of the scene by e.g. scanning through the workspace row by row.

To summarize, table 6 below lists all functions, per feedback type, available in the application just described.

Table 6. All functions available in the application.

<table>
<thead>
<tr>
<th>Type</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>The workspace is seen from the same fixed angle by both users</td>
</tr>
<tr>
<td>Haptic</td>
<td>Feeling the bottom plate, cubes and space between cubes</td>
</tr>
<tr>
<td></td>
<td>Feeling differences in friction between cubes and bottom plate</td>
</tr>
<tr>
<td>Audio</td>
<td>Fall sound – heard constantly from the falling cube</td>
</tr>
<tr>
<td></td>
<td>Touch down sound – heard when a cube has just landed</td>
</tr>
<tr>
<td></td>
<td>Scan sound – heard when scanning with the proxy over piles</td>
</tr>
<tr>
<td></td>
<td>Empty scan sound – heard when the proxy is moved over an empty cell</td>
</tr>
<tr>
<td></td>
<td>Ending sound – heard when all 10 empty cells have been filled out</td>
</tr>
</tbody>
</table>

6.3 The experiment

Based on the results from the studies on collaboration between sighted and non-sighted users the new experiment continued the investigation of influences of modality combinations on collaboration and communication, but with sighted users. In this case within subjects experiments were performed with the conditions visual/haptic, visual/audio and visual/haptic/audio counter balanced. The work in this study was guided by the following general research question:

How is task performance affected and employed strategies influenced in a virtual 3D collaborative multimodal environment, depending on the modalities currently supported?

The research hypotheses, derived from the overall research questions, were the following (see section 6.4.1 for detailed information about what defines an error):

H1: *In the condition with vision, touch and hearing supported, the task will be performed significantly faster than in the visual/haptic and the visual/audio conditions.*

H2: *In the condition with vision, touch and hearing supported the number of errors made will be significantly lower than in the visual/haptic and the visual/audio conditions.*

Apart from evaluating the above hypotheses a main focus, again, was on performing qualitative analyses of observation data. This approach is more thoroughly explained in chapter 3.
6.3.1 Subjects

26 university students were recruited for this experiment, all working in pairs. They were all recruited from the first two years of the Computer Science program at KTH and had no prior experience of haptic feedback. To promote collaboration 13 of the students accepting the invitation got to choose the person, from the Computer Science program, they wanted to work with during the experiment.

6.3.2 Apparatus

As before, a DELL Precision PC with two dual core processors was used to run the program. Two PHANTOM Desktop devices, each connected directly to the computer, were used as haptic equipment for the two participants. Two speakers were used for the audio feedback. Two screens were also connected to the computer, through a switch. Each user also had their own keyboard and a mouse was used by the test leader to load assignments.

Since the users sat in two adjacent rooms (see section 6.3.3 below) two laptops were used, in parallel with the equipment necessary for running and interacting with the program, to run Skype during the sessions. Last, a video camera was used to capture the screen and the discussions between the participants, in one of the rooms.

6.3.3 Setting

The setting used in the experiment is illustrated in figure 22 below.

![Figure 22. The setting used in the experiments.](image)

During the entire experiment session the users sat in two adjacent rooms. The reason for this was mainly that we wanted to see how consensus could be reached by the participants if they were not able to point directly at the screen to illustrate positions and strategies. To make collaboration possible at all times during the experiment a Skype connection was always running on the two laptops placed beside the participants (see
To enable simultaneous interaction both users had both a keyboard and a haptic device on the table in front of them.

The video camera was held steady during the entire experiment and placed in the room with the computer. The image on page 98 shows an example image captured from the camera and also illustrates the working situation.

### 6.3.4 Procedure and tasks

Since the experiment was based on the within group design the same task was solved three times by each group under the following conditions:

- Visual/haptic
- Visual/audio
- Visual/haptic/audio

The functions and feedback types available are thoroughly described in section 6.2 about the application tested. The conditions were switched by turning on/off the speakers and haptic feedback respectively. Note that the haptic devices were used as pointing devices even in the visual/audio condition even though haptic feedback was not enabled.

The experiment consisted of 8 main parts: (1) information to participants, (2) training task condition 1, (3) task in condition 1, (4) training task condition 2, (5) task in condition 2, (6) training task condition 3, (7) task in condition 3, (8) interview with both participants.

As usual, the experiments started out with the researcher providing introductory information about the experiment and how the application worked, without revealing the goals of the research or which conditions to test. After this, the Skype connection was initialized and one of the users was escorted to the other room.

When both participants were seated at their respective work stations the application was loaded for the first time, in the first condition. They now got the opportunity to test all functions available in this condition during a training session lasting about 15 minutes. Except for training on moving the falling cubes with the arrow keys the participants tried out the haptic and/or audio functions depending on which condition the group started out with.

After the first training session the application was closed and the participants were provided with the following task, verbally:

*The application will now be restarted with a random configuration of 150 cubes on the red bottom plate. On an unknown number of cells there is no pile of cubes, meaning that the red bottom plate can be seen from above. You should use the information available to collaborate in finding these empty cells and filling them with cubes falling from above. You will hear an ending sound when all empty cells have been filled out.*

The application was then restarted and the participants worked together to solve the first task. When the ending sound was heard (all the 10 empty cells identified and filled out) the application was closed again. This same procedure – training, task – was repeated.
three times (one per condition). The training was, of course, only based on the types of feedback that were available in the particular condition.

The last part of the experiment sessions consisted of a short semi-structured interview aimed at eliciting opinions about the interface and gaining insights into how the participant thought that the change of conditions influenced their communication and overall task solving strategies. The following questions were asked:

- In which of the tasks was it easiest to solve the task together?
  - Why do you think that this task was the easiest?
- Which type of function did you find most usable?
  - Why was it more usable?
- Which types of feedback were easy to use?
- Which types of feedback were hard to use?
- How did the change in conditions affect the way you communicated with each other?
- How did the change in conditions affect your means of solving the tasks?

6.4 Analyses

The qualitative analyses were, as in the previous studies, based on video recordings and interviews. The general procedure followed for analyzing the recordings is the same as in the other studies and explained in section 3.1 about analyzing dialogues. The categories relevant for this thesis are listed in the next two sub-sections.

The interviews were also transcribed in their entirety. After transcription, the different parts of the interviews were summarized and sorted under the respective questions asked. This was done for each group. Answers related to the different feedback combinations were compared in a last iteration.

6.4.1 Quantitative measures

- **Time to complete task**
  The time it took for the participants to solve the test task. The time was taken three times per group – one per experiment condition. The timer started when the application was loaded and stopped when the ending sound was heard. This measure was subject to statistical analysis and corresponds with hypothesis H1. This measure is meant to shed light on if access to more senses decreases task completion time. Potentially, it can also tell us which one of the audio and haptic modalities is the most efficient to use in the type of sorting and scanning task used in this experiment.

- **Number of errors made**
  The number of errors made during the task solving process, calculated three times per group (one per experiment condition). In this case an error equals misplacement – when a cube lands on a spot which is obviously not intended by the one controlling the cube. Thus, this measure is dependent on the verbal
communication between the participants since some kind of confirmation of the
error is needed, but still serves as a good indication of efficiency of the interaction
in the different conditions used. This measure was subject to statistical analysis
and corresponds with hypothesis H2.

- **Number of result checks**
The number of times the participants checked the result of a cube placement, to
see if an empty cell was indeed filled out. This measure also shed light on the
efficiency of the interaction in the different conditions and can e.g. tell us if result
checks are less frequent in the condition with all three senses supported. This
measure did not correspond with a pre-defined hypothesis but was still subject to
statistical analysis.

- **Number of double checks**
The number of times a participant double checked the occurrence of an empty cell
found by the peer. As for most quantitative measures the number of double checks
performed also shed light on the efficiency of the interaction in the different
conditions. This measure can tell us if the perceived need to double check the
partner’s work, or to ask the peer for a confirmation, decreases when more
modalities can be used. As for the result check measure, this measure was subject
to statistical analysis, but did not correspond with a hypothesis.

6.4.2 **Qualitative measures**

The following list contains an overview of the most central categories used in the
qualitative analysis.

- **Division of labor**
The means by which the participants divide the work between each other. One of
the most important aspects considered here is the difference in division of labor
between the visual/haptic/audio condition and the other two conditions. Results
from earlier studies presented in this thesis (Moll et al., 2010) suggest that
participants tend to work more concurrently when having access to more
modalities.

- **Scanning**
The means by which the participants scan the environment in search for empty
spaces, in the different conditions. Important aspects are how one can use haptic
and audio feedback, respectively, in the scanning phase and how the strategies
differ in the visual/haptic/audio condition compared to the other conditions.

- **Establishing position**
The means by which the participants established the position of the empty cells
found and what methods they used for placing the falling cubes. As for scanning,
both the methods used with haptics and audio and the differences between conditions are important elements to consider here.

- **Grounding**
The means by which the participants establish a common ground about the layout of the workspace, the result of particular actions and the controlling of the falling cube. The amount of grounding needed in the different conditions is of special importance here.

- **Location awareness**
The means by which the participants stay aware of each other’s whereabouts during the task solving process. Discussions of content not visible to the eye are of special interest here.

### 6.5 Main results

The main results from the experiment will be provided in this section, both from the quantitative and qualitative analyses. As usual, dialogue excerpts play a central role in the discussion. In all of these excerpts the two participants are denoted P1 and P2. For more examples and results the reader should consult the resulting article (Moll et al., 2013).

#### 6.5.1 Quantitative results

All the quantitative results from the experiment can be seen in table 7 below. Friedman’s test was used for the quantitative analysis, since ANOVA could not be used due to the data not being normally distributed. A Wilcoxon signed rank test with a Bonferroni correction applied was used for the post-hoc tests.

<table>
<thead>
<tr>
<th>Table 7. The quantitative results gained from the second experiment (Moll et al., 2013).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio/Haptic/Visual feedback</td>
</tr>
<tr>
<td>Performance (sek.) (n=13, (\chi^2(2)=25.63^*)) M=250 SD=47.7</td>
</tr>
<tr>
<td>Misplacements (n=13, (\chi^2(2)=18.89^*)) M=1.23 SD=0.73</td>
</tr>
<tr>
<td>Double checking (n=13, (\chi^2(2)=26.00^*)) M=1.38 SD=0.96</td>
</tr>
<tr>
<td>Checking of result (n=13, (\chi^2(2)=21.64^*)) M=2.08 SD=0.95</td>
</tr>
</tbody>
</table>

* = significant at 99.9% level

There are several points worth noting among the above results. First of all the result shows a significant difference between the three-dimensional visual/haptic, visual/audio
and visual/haptic/audio conditions regarding all the measures that were subject to analysis. The significances were all on the 99.9% level.

We can also see from table 7 that the standard deviations were the smallest in the visual/haptic/audio condition. Thus, the results are most secure in that condition and this indicates that having access to more senses increases the precision of work, overall.

When it comes to the task performance measure post-hoc tests revealed that task performance was significantly faster in the visual/haptic/audio condition than in the visual/haptic (Z=3.180, p=0.001) and visual/audio (Z=3.110, p=0.005) conditions. Task performance in the visual/audio condition was also shown, not surprisingly, to be significantly faster than in the visual/haptic condition (Z=3.181, p=0.001). These results are very clear and confirm hypothesis H1 – task performance is significantly faster in the condition with all senses supported.

When the number of errors/misplacements is concerned post-hoc tests revealed that the number of errors was significantly lower in the visual/haptic/audio condition than in the visual/haptic (Z=3.078, p<0.005) and the visual/audio condition (Z=2.924, p<0.005). The number of errors made was also significantly lower in the visual/audio condition than in the visual/haptic condition (Z=2.507, p<0.05). These results are also clear and confirm hypothesis H2 – the number of errors made is significantly lower in the condition with all senses supported. This supports the claim that the quality and precision of the work performed is positively affected by the support for additional senses.

Regarding the number of double checks performed the post-hoc tests revealed that they were significantly less frequent in the visual/haptic/audio condition than in the visual/haptic (Z=3.190, p=0.001) and visual/audio condition (Z=3.204, p=0.001). The number of double checks was also significantly lower in the visual/audio condition than in the visual/haptic condition (Z=3.195, p=0.001). This is a strong indication that the participants felt more secure of the interaction in the visual/haptic/audio condition than in the other conditions. Since the times to perform tasks were also shortest and the number of errors were lowest in the visual/haptic/audio condition the fact that very few double checks occurred (M=1.38) did not seem to affect task performance in a negative way. This is an important conclusion when it comes to the potential of using more senses for the types of tasks used in the experiment.

The last quantitative measure concerned the number of result checks. Even in this case post-hoc tests revealed that this number was significantly lower in the visual/haptic/audio condition than in the visual/haptic (Z=3.216, p=0.001) and visual/audio condition (Z=3.089, p<0.005). The number of result checks was also significantly lower in the visual/audio condition than in the visual/haptic condition (Z=2.547, p<0.05). This is a strong indication of that the participants felt more secure of their own interaction in the visual/haptic/audio condition than in the other conditions. As was the case for the number of double checks, this result did not cause worse task performance.

All the quantitative results are important and they reveal interesting facts about the potential of adding support for more senses. However, they do not reveal any concrete facts about the influences that modality combinations have on the collaboration and especially the communication between the participants. This is why the qualitative analysis, which will be discussed next, plays a central role in this study.
6.5.2 Qualitative results

As it turned out, the qualitative results gone through in this section yielded many interesting insights into how the combination of modalities influenced the collaboration and the task solving strategies employed. The qualitative results gained also, in large, explain the quantitative results gone through earlier.

6.5.2.1 Typical task solving strategies

In this section, typical strategies in the different conditions will be briefly discussed due to their important implications for the time to complete the tasks.

In the visual/haptic condition the typical strategy used for scanning for empty cells was to go through each row or column systematically from one side to the other. The participants were often seen following the piles along each row up and down again to make sure no space was missed. Even though most of the groups started out by going down feeling behind piles in a random fashion almost everyone gave up that strategy in favor of the more systematic one. Example 12 below illustrates a typical situation, occurring in the visual/haptic condition.

Example 12:

P1: No, we have to do this systematically
P2: Yes, let’s go from each side. We start by going row by row
     [P2 starts at the left side going right row by row]
     [P1 starts at the front row going forward row by row]
     [P1 finds an empty cell on the second row after a few seconds]

P1: Here, we have one
    [P1 moves his proxy up and down to show the position]

When trying to place the falling cube over the empty cell to be filled, one common strategy in the visual/haptic condition was to start at the front row and approach the found empty space row by row until the empty space was felt. This strategy was time consuming, but enabled the participants to calculate how many steps from the front row the falling cube needed to be moved. Another common strategy was to move the proxies upwards from an empty space, aiming at the falling cube from below, using the fact that they would get hit by the falling cube if it was over the right cell.

In the visual/audio condition the strategies used were somewhat different. The most common strategy used when scanning for empty spaces in this condition was to make fast sweeping movements randomly over the entire workspace listening for the empty scan sound. When the sound was heard more fine-tuned movements over the identified area were performed in order to localize the exact position. When trying to position the falling cube, in order for it to fall down into the empty space found, the most common strategy, again, was to start at the front row and listen to the cliques while moving forward slowly approaching the empty scan sound. Aiming with the proxy from below was also quite common. However, more misplacements occurred with this strategy in the visual/audio condition than in the visual/haptic one.
In the visual/haptic/audio condition it was especially interesting to note how the different senses complemented each other. When scanning for empty cells the sound was almost always used in the same way as in the visual/audio condition – rapid sweeping movements. This is not surprising, since it is a lot harder and time consuming to scan a complex environment using the sense of touch (Kahol and Panchanathan, 2008). When the empty scan sound was heard the haptic feedback was often used to establish the exact position of the empty cell. When trying to position the falling cube over the empty cell to be filled out, the most common method was to aim with the proxy from below, trying to get hit from above. Example 13 illustrates a typical discussion in the visual/haptic/audio condition. This example will be discussed more later on in this section.

Example 13:

P1: Here is a hole [an empty cell] ... here!  
[P1 moves her proxy up and down over the empty cell to indicate its position]
P2: Yes  
[For some reason the next falling cube is not moved by any one]
P2: There is still a hole here  
[P2 moves back and forth over the cell and P1 also joins him]
P1: Yes  
P2: I have to go down and feel here  
[P2 goes down on a place where the cell is heard]
P1: Yes  
P2: Oh, it's there!  
[P2 moves his proxy upwards from the cell and aims for the falling cube]  
[The strategy works and he manages to fill the empty cell]

Example 13 and the above discussions illustrate important points about the benefit of working with three senses. When presented with both haptic and audio feedback the participants chose to scan with the sound and establish position with the haptic feedback. The senses complement each other in that it is possible to make use of audio for a quick overview during the scanning face and the haptic feedback to make a more exact judgment of position.

6.5.2.2 Influences on communication

In the study presented here haptic and audio guiding functions were not evaluated, but nevertheless the combination of modalities currently supported was shown to influence the communication between the participants. The communication between the participants also, most likely, explains parts of the quantitative findings presented earlier.

One clear difference between the experiment conditions was that less discussion occurred in the visual/haptic/audio condition than in the other conditions. Even though no statistical analysis was performed on the amount and/or length of utterances used during the task solving process this result was clear from the observations. A reasonable cause for this can be found when looking at why the participants talked to each other and in what situations they did it.
In the visual/haptic groups it was clear that the hardest part of the work was to use the haptic feedback to scan for empty spaces (see section 6.5.2.1 above). In most cases it took the participants a long time to find the empty spaces and this is most probably the reason why they spoke a lot when working in that condition, especially when grounding the impression that an empty space was indeed found at a particular spot. The participants communicated about almost all empty spaces found in this condition and most of them confirmed in the interviews that they discussed a lot and double checked each other just because they wanted to be extra sure in a situation when the empty spaces were hard to find. This also, in general, explains the quantitative results on double checking and result checking. In the visual/haptic groups the participants also talked a lot about strategies for solving the task as efficiently as possible. All in all, the participants felt forced to communicate in the visual/haptic condition a lot due to the difficulty of using haptic feedback to get an overview of a complex environment. Example 12 given earlier shows a typical discussion in almost all groups.

In the visual/audio condition discussions among the participants were also frequent. Even in this case the participants spoke a lot about what to do next and quite often they focused their discussions and concentration on one empty space at a time. Results from interviews indicate that this was much due to the difficulty of establishing the exact position of the empty spaces just by using the sense of hearing. Focusing on one empty cell at a time, like in the visual/haptic condition, of course affects the task completion time in a negative way.

In the visual/haptic/audio condition the situation was quite different, regarding the amount and focus of discussions. First of all, the participants talked a lot less in this condition than in the other two and this is most probably due to the fact that they were more confident with their own interaction when working in that condition. They had access to information from more sensory modalities complementing each other, as discussed before. This could also be confirmed when considering the focus of the discussions between the participants in this condition – very few utterances concerned the feedback gained and/or uncertainties in the sensations. Discussion about strategies to solve the task almost only occurred at the outset of the task. When working in the condition with three senses supported the participants almost only talked when they had found an empty cell and needed to coordinate the use of the next falling cube. All in all, it was clear that the combination of modalities currently supported influence the communication between the participants both when it came to the focus of the discussions and the number of utterances. To conclude this section, example 14 below illustrates a typical discussion in the visual/haptic groups.

Example 14:

P1: I think there is another one, right beside [an earlier filled empty cell]
    [P1 feels on the spot a little bit longer]
P1: Or?
    [P2 comes to the same place and feels for a while]
P2: No, there’s not
    [P1 continues to feel a few seconds]
P1: No, there wasn’t
6.5.2.3 About referencing
Even though no clear difference could be seen between the experiment conditions regarding the means of referencing and establishing common ground in the interface, some points are still worth noting. Let us use example 15 below as an illustrative starting point.

Example 15:

P1: *Ok, check here!*
   [P1 moves her proxy to show the place and P2 joins]

P2: *Yes*

P1: *I think it’s behind this somewhere*
   [P1 points to a high pile]

P2: *Yes, but it’s not just…*

P1: *It’s several, right?*

P2: *Yes*

P1: *It’s a triangle there [three empty cells in an angle, that were felt]…*
   [It is silent for a few seconds]

P2: *Yes, it must be*

The above example is interesting in many respects. Most importantly, the participants discuss the layout of the scene effortlessly despite the fact that none of them can actually see the situation. They also manage to reach consensus about the arrangement of empty cells in this particular case. The haptic feedback that both users get makes this establishment of common ground possible and straightforward.

Through a more detailed study of example 15 we can also see how haptic feedback enables deictic referencing by both participants. Almost in every utterance a deictic reference to a place not seen by the participants, like “here” and “it”, is made. This is especially important for the quality of the communication and the efficiency or work, since explaining the current situation without using deictic terms would be very cumbersome in this visually demanding interface.

Although it does not concern the haptic feedback it is also worth noting that deictic references are also made through pointing with the proxy at e.g. piles close to a target area. This method of directing the peer’s attention to a particular area was used frequently in all conditions.

Even though example 15 is taken from work in the visual/haptic condition similar discussions and means of establishing common ground occurred in the other conditions as well. Of course, the sounds were used to establish consensus in the visual/audio condition.

A last important point to make is that the example clearly shows that the participants are aware of each other’s whereabouts in the environment. Example 15 is representative in this respect, in that the participants often discussed how the scene looked at a particular spot without actually seeing each other’s proxies.
6.5.2.4 Influences on collaboration

The modality combinations did not only influence the focus of communication between the participants but also the collaboration as a whole. That the specific task solving strategies regarding scanning for and establishing the exact position of identified empty cells influenced the collaboration, has already been discussed. More subtle influences were however, also found.

Not only did the participants, mostly, focus their discussions on one empty cell at a time, regardless of who found it, in the visual/haptic and visual/audio conditions. They also often worked together on establishing the exact position of it and filling it out. In a typical case one of the participants announced that he/she had found an empty cell. Almost immediately the peer went to the same place to confirm that the found cell was indeed empty. After they agreed on the fact that there was an empty cell they often collaborated on calculating the position of the cell and filling it out with the next falling cube. This collaboration around the empty cells occurred regardless of whether both participants had found empty cells in parallel or not. To collaborate around each empty cell like this of course has effects on the overall task performance.

In the visual/haptic/audio condition the situation, regarding collaboration, was totally different. Example 16 below illustrates one of the usual approaches.

Example 16:

P1: Here is the next hole [empty cell]
P2: Yes, fill it
   [P2 now continues to search further back in the environment]
   [P1 takes the next falling cube and positions it above his proxy]
P2: I think I have several holes here
   [P2 is now seen moving his proxy back and forth around an area in the back, as a way of showing P1 which area he is referring to.]
   [After a few seconds P2 is joining P1 to help him position the cube, by also holding up his proxy above the empty cell]
P2: Yes, that’s good
P1: Yes, it’s right

In this case, the participants work a lot more in parallel. In the first phase of the dialogue P2 just acknowledges the fact that P1 has just found an empty cell. P2 then continues searching while P1 positions the next falling cube over the empty cell found. Later on, when P2 has found several empty cells, and P1 is done with his empty cell, both participants team up at P2’s location. The important thing to note here, which was clear from all groups when working in the visual/haptic/audio condition, was that the participants worked concurrently solving their own sub-tasks as long as both participants had their own specific task to solve.
6.6 Lessons learned

As can be seen from the dialogue excerpts and the discussion in the result section 6.5 many interesting results were gathered, relating to how the modality combinations influenced the collaboration between the participants. Above all, it is clear that the collaboration is eased, in a number of respects, when the participants have access to all three senses. This section will discuss the most important results more deeply.

6.6.1 About scanning and placing strategies

Although the task completion times and the means of scanning and placing changed with the conditions, the participants managed to solve the proposed task in each condition. When comparing the conditions with two senses supported (visual/haptic and visual/audio) the results are apparent – the audio feedback is more suitable for the initial rapid search and the haptic feedback is more suitable for establishing exact positions and placing the falling cube. The inefficiency of the haptic feedback during the scanning phase is, as stated earlier, due to the difficulty of establishing an overview of a complex environment using solely touch feedback (Kahol and Panchanathan, 2008). The situation becomes particularly challenging and time consuming when using a one-point interaction device for the touch feedback (Lederman and Klatzky, 2004). In the visual/audio condition, on the other hand, the scanning phase was all about moving rapidly over the interface listening for a particular sound (the empty scan sound). Thus, when comparing the two methods available for scanning it is not surprising that scanning is more efficient in the visual/audio condition.

However, when establishing exact positions in the interface the audio feedback falls short. If no haptics is available the proxy just becomes a 3D pointer. The scan sounds reveal how many cubes are currently under the proxy and the empty scan sound tells the user that the proxy is over an empty cell, but the graphical interface still poses a challenge due to the perspective used. This was confirmed by many participants during the interviews. You hear that you have an empty cell at a particular spot, but you can only judge its approximate location. In the visual/haptic case the situation was completely the opposite – due to the ability to feel everything in the interface, more exact judgments about position could be made (see example 13 on page 110 for an illustration). This may in part by explained by the meta study performed by Hale and Stanney (2004). They draw the conclusion that haptic feedback can be a considerable aid when judging object sizes and distances to and between virtual objects. In a visually demanding interface such as the one discussed in this chapter, this additional aid seem to make a big difference. When an empty cell has been reached the establishment of the exact position can be judged by a combination of the very limited visual cues and our motor abilities.

In the placing phase haptic feedback was also more efficient to use. Even though not all of the groups realized that their proxies could be hit from above by the falling cube (every part of the interface could be felt), those who did almost always placed the falling cube at the first try. In the visual/audio condition this very physical cue was not there and for the same reasons as discussed above the positions of the raised proxies were probably less exact.
All this being said, it is quite natural that the task completion times were significantly faster in the visual/audio condition than in the visual/haptic one. Even though the haptic only version seemed to be more appropriate when establishing positions and placing falling cubes, the inefficiency of using haptic feedback for scanning made all the difference.

The most interesting result emerges when merging the two environments into the condition supporting all three senses. The quantitative results were clear – task completion times were significantly faster in the visual/haptic/audio condition than in any of the other conditions. The important question to consider is why this result was gained despite the fact that the sound was so superior, in comparison to the haptic one, in the scanning phase. As discussed briefly in the result section (6.5) the senses complemented each other in this condition – the audio feedback could be used to find new empty cells rapidly (the scan phase) and the haptic feedback was used for the more detailed work of establishing exact position of empty cells and placing falling cubes in these cells. When adding audio feedback to the haptic interface, i.e. adding a much more effective means of scanning, it does not come as a surprise that task completion times will drop considerably. This result has also been found in a different setting in the study described in chapter 5 (Huang et al., 2012). It is more interesting, however, to consider the fact that task completion times drop significantly when haptic feedback is added to the audio interface – the haptic feedback provides a significant added value.

The discussion in this section has up to this point only concerned individual interaction with the environment. However, the results on task performance relates to two users collaborating. The particular strategies for scanning and placing made possible by haptic and audio feedback respectively, can explain parts of the results on task performance. The collaborative tactics, however, enforced by the scanning and placing strategies made possible in the respective conditions, can give a deeper explanation. Some of these collaborative tactics will be discussed in the next few sections.

6.6.2 Influences on group dynamics

The most important results from this study relates to how the combination of modalities influences the collaboration between and the work strategies employed by the participants. The examples brought up to discussion in the result section show that the participants needed to adjust their means of communicating and collaborating depending on the modalities currently supported. Above all, less communication occurred in the condition with three senses supported and there was considerably more concurrent work in that condition. In the earlier studies discussed in chapter 4 and 5 pure communicative functions based on haptic and audio feedback were tested with pairs of users collaborating in virtual environments. In the study presented here, no such functions were implemented – haptic feedback was solely used for feeling interface elements and audio feedback was used for listening to the falling cube and one’s own interaction in the environment. Thus, the results reported here, regarding influences on collaboration and communication, do not stem from functions especially designed to be used for communication. In this respect the question studied in this chapter is more general – the specific modality combinations and feedback types are in focus and not particular
communicative functions. In this respect the study is quite unique, even though some of the results might be expected.

That the changes between conditions would have influences on the group dynamics in the pairs of participants was not at all expected from the beginning, at least not when it comes to the magnitude of the effects. In the case with three senses supported the strategy applied by all pairs was to divide the work of scanning and placing between them. In the case of the other conditions the strategy was to divide the work of scanning, but work closely together while establishing exact positions and placing falling cubes. These strategies were also confirmed by the participants from most groups in the interviews concluding the experiment sessions. It is obvious, from both the observations and interviews, that the participants were more confident with their own interaction in the condition with three senses supported and it feels safe to argue that this is due to the ability of haptics and audio to complement each other in that case.

The changes in experiment conditions also had a great impact on the means of communicating – a lot less was said during the interaction in the visual/haptic/audio condition than in the other two conditions. It is interesting to relate this result to the concept of grounding discussed by Clark and Brennan (1991). Even though the participants taking part in the study may not have thought about it consciously, it is obvious that there was more need for grounding in the two conditions where the available modalities (haptics or audio) were not suitable for scanning and placing respectively. This is another way of saying that the participants need to work harder to establish and maintain a common ground and that the felt need of establishing a common ground is more pronounced when the task is more challenging. In the visual/haptic and visual/audio conditions grounding activities were performed on many different levels, like the result of a particular action, the layout of the interface and the controlling of the falling cube. In the case with three senses grounding mostly occurred when discussing who should control the falling cube.

Roque and Traum (2008) proposed a taxonomy to be used when discussing different means of communicating in order to establish common ground (see section 2.2.1.2). When related to the results presented here one can conclude that submit (like “I have found a hole. I will take the next falling cube”) followed directly by an acknowledgement (like “ok”, yes”,...) was the, by far, most utilized means of reaching an agreement in the case with three senses supported. During the work in the other conditions there were often more utterances of the types re-submit and request repair before a common ground could be reached. Note that the taxonomy proposed by Roque and Traum (2008) was never a part of the analysis in this study and this is why it has not received much attention here. Nevertheless, it is useful when studying how the means of communicating and establishing common ground changed between the conditions.

The cost of grounding is also worth relating to here. According to Clark and Brennan (1991) communicating partners not only use different mediums and tactics when grounding but they also tend to weigh the cost of grounding to the current topic and the current medium. This phenomenon is called “least collaborative effort”. From the results of the experiment, it is clear that the participants invested more effort in grounding activities in the visual/audio and visual/haptic conditions than in the visual/haptic/audio condition. Since the topic of discussions (about the tasks of scanning for and filling out empty cells) was always the same in all the groups, the participants must have changed
their means of grounding because of the changes in the medium. Clark and Brennan (1991) mostly discuss the costs of using different kinds of spoken and written media, but the same discussion could also be applied in the case of the conditions used in the experiment presented here. When both audio and haptics are present there is a good amount of redundancy and the participants always have access to a sense which is suitable for a particular subtask (scanning, placing,...). The cost associated with grounding in this condition is the lowest and, as stated earlier, this is obviously recognized by the participants. On the other hand, the cost associated with grounding in the other conditions is higher and this is reflected by the participants talking more.

The issues of uncertainty and aiding are also important parts of the result. When combining the observational data and the quantitative data from the experiment it becomes apparent that the level of uncertainty that the participants feel about their own and the other’s actions is lowest in the condition with three senses and highest in the condition with only haptic feedback supported. The need to check one’s own and the other’s actions (i.e. result checks and double checks) was much higher in conditions where only one sense was supported. This result is important, since uncertainty is generally detrimental in collaborative situations – the results of the experiment show that the conditions in which the participants felt most uncertain was the ones associated with the highest cost for grounding. Thus, even when it comes to the level of uncertainty the condition with three senses was shown to be the best.

6.6.3 Sound and haptic feedback for emptiness

Two studies were actually performed with the interface developed for this part of the thesis, and the focus of this chapter has up until now been on the second one. The reason for conducting two studies was the design of the “empty cell” in the audio implementation in the first version of the interface, which directly affected task performance. Statistical analyses from the first study, from here on called pre-study, show that tasks are performed significantly faster in a visual/haptic/audio condition than in a visual/haptic one and that tasks are performed significantly faster in a visual/audio condition than in a visual/haptic one. There were, however no significant differences between the visual/audio and the visual/haptic/audio condition. The pre-study was never published, but it will be covered here to illustrate the importance of the audio design.

Two of the sound functions described in section 6.2 were added after the pre-study – the empty scan sound and the ending sound. In the application tested in the pre-study no sound was heard when the proxy entered an empty cell. The implications of the change of representation of an empty cell – a special sound for emptiness instead of silence – proved to be important. The fact is that the changes made, before the experiment presented in earlier sections was carried out, had a pronounced effect on the task performance. Furthermore, with the addition of the ending sound the quantitative task completion time measured the time it took to fill out all the empty cells present and not the time it took to find as many empty cells as possible (without knowing if you had missed any), as was the case in the pre-study.

When “designing for nothing” there are always two possibilities – (1) let the user experience emptiness directly (i.e. no feedback) or (2) let the user experience feedback
which in turn represents emptiness. In the application used in the pre-study the haptic environment had a haptic feedback representation for emptiness (lower friction) while the sound environment let the user experience emptiness directly by silence (lack of scan sound). These design choices are not obvious and the choices made did have significant effects on the task performance (but not so much on the collaboration). Even though the effects of changing the sound model did not have a clear impact on the collaboration it is worth elaborating on the topic some more – there is a clear need for future research here.

In the pre-study, silence represented empty cells, where no cubes were placed, and this may seem as the most natural design. One of the developers of the sound system, Sten-Olof Hellström, had also been involved in an artist-driven project where silent “bubbles” in mid-air were created in a room full of sounds (Rosén et al., 2003). In their case the sound design could be used to effectively convey a sense of emptiness and this fact inspired the design of emptiness in the first version of the sound system used in the pre-study.

When interacting with the sound system used in the pre-study the scanning phase was all about moving slowly from pile to pile listening to the scan sounds indicating how many cubes were currently under the proxy. When no sound was heard when moving from one cell to another there was an indication that an empty cell had been found. The core problem here was that the participants experienced difficulties in deciding if there was an empty cell at a particular place or not – it was quite easy to miss an empty cell even though the proxy passed over it. The problem stems from the fact that it is much harder to search for lack of feedback than to search for actual feedback, especially in the interface evaluated. When moving with the proxy over a pile a series of cliques tells the user how many cubes are currently under the proxy. This information could be vital to task completion since it, at least to some degree, can aid the 3D perception of the interface – the user can hear where in the interface a certain high pile is located, for instance. If the user moves too fast over the piles an empty space can easily be missed, especially in the far back of the interface. When an empty cell was found (by attending to the lack of feedback) the participants often moved slowly back and forth just do make sure that they actually had found an empty cell. Thus, the process of scanning for empty cells was most about the participants’ abilities to move slowly and at a constant pace – the participants’ motor abilities played a central part. This is problematic, since the idea with the experiment was to shed light on strategies for scanning and not to examine how long it took to actually realize that an empty space had been found during the scanning phase.

An alternative representation of the empty cell was tested in the experiment. In this case a special kind of sound (empty scan sound) was heard every time a user passed over an empty cell. Thus, in the experiment there was a special kind of audio feedback which represented emptiness. Naturally, the participants experienced no difficulties in realizing if they had found an empty cell or not during the scanning phase and at the same time they still needed to move a few times back and forth to determine the exact position of the empty cell (in the visual/audio condition, at least).

When looking at the quantitative results on task performance gained from the pre-study and experiment respectively an important fact to consider is that the task was solved significantly faster in the visual/haptic/audio condition than in the other conditions in the experiment, but not in the pre-study. In the pre-study the results showed that the
task was performed faster in the visual/haptic/audio condition than in the visual/audio condition, but significance at the 95% level was not reached. The reason why significance was reached in the experiment, as thoroughly discussed in section 6.6.1, was most probably that the haptic feedback made a contribution in the phases of establishing positions and placing. This added value must have been present in the visual/haptic/audio condition in the pre-study as well, but it was blocked out – the difficulty of knowing if an empty cell had been found or not in the visual/haptic/audio condition in the pre-study reduced the added value of the haptic feedback. This result shows how important the design of every function is – the change of representation of the empty cell was needed to show the value of haptic feedback when added to a visual/audio interface.

The design decision regarding the haptic representation of emptiness was not changed after the pre-study, but even in this case there were two possibilities. Instead of representing the empty cell as difference in friction it could also be represented as a physical hole (i.e. the proxy falls through the workspace when an empty cell is found). This alternative solution is not at all logical, mostly because of the visual interface, and above all the navigation would be severely hindered if the users fell through the workspace at the target areas. The important issue here is that the type of representation used for the empty cell was the same for the haptic and audio feedback in the experiment in that both utilized a special kind of feedback representing emptiness. Even though the means of interacting in the haptic and audio interfaces differ considerably, the fact that they share the same design principles make it easier to compare them.
7 Discussion of contributions

In this section the contributions from the thesis project will be further discussed and elaborated on. Results from the different studies will be compared in order to derive more general conclusions. Before moving on I want to remind the reader of the main research question on which this thesis is based:

*How do changes in modality combinations influence employed work strategies, communication during task solving and the task efficiency in collaborative multimodal virtual environments?*

The studies discussed in chapters 4-6 all address this question from different angles.

7.1 The combination of modalities matters

When the work with the thesis studies started the number of previous studies explicitly comparing modality combinations in collaborative interfaces was surprisingly small. A few examples are a virtual air hockey game (Nam et al., 2008) and two environments for joint handling of shared objects (Sallnäs, 2004). Nam et al. (2008) studied the effects of modality combinations on user performance. They used the conditions visual, visual/haptic, visual/audio/haptic in their study and concluded that access to more modalities aided the users in a number of different respects. Sallnäs (2004) only studied the conditions visual, visual/haptic but the conclusion that haptic feedback has a pronounced effect on the task performance time and the feeling of presence in the interfaces was reached in those studies as well.

In the study described in chapter 4 no modality combinations were compared against each other since audio was not used and visually impaired users were an important part of the target user group. In chapter 5, however, the combinations visual/haptic and visual/haptic/audio were compared and it was shown that the addition of audio feedback had a significant effect on task performance measured in time to complete a task. The unique area studied in this case was collaboration between sighted and non-sighted persons. The literature review performed during the thesis work gave no indication that task performance has been compared between modality combinations in collaborative interfaces in that setting before. The review showed, however, that there are numerous studies in which comparisons have been made between modality combinations in single user interfaces both for visually impaired and sighted users. Some of these are Yu and Brewster’s (2003) work on exploration of bar charts and the studies referenced in the meta-study performed by Burke et al. (2006). Another key researcher when it comes to the effect of different modalities on user performance is Charles Spence who has performed several studies on cross-modality relations and attention. The study of his which is most important for the arguments presented in this thesis (Lee and Spence, 2008), described thoroughly in section 2.1.3, is about the pronounced benefit of using more modalities than vision when coping with demanding dual-task scenarios.
When testing collaborative applications the situation is more complex and it is not an easy task to fully explain the results gained. When judging the results on performance in a collaborative interface one cannot just look at individual performance but also on joint performance. This is one of the reasons why methods were triangulated in this thesis – both quantitative performance measures and qualitative data about the collaboration and communication were used in all studies performed. The performance measures showed that task completion times decreased significantly as the number of available modalities increased but these measures alone could not explain why. The in depth qualitative analysis of how the communication changed depending on the combination of modalities shed light on the reasons behind the quantitative results. Looking at performance measures and interpreting them as indicators of the effect that the change in modality combination causes has been done several times before, especially in single user settings, but the triangulation gained from studying dialogues and collaboration in depth is rare. One example of such a triangulation (a qualitative analysis explaining quantitative data) can be found in the study on collaboration between sighted and non-sighted students, presented in chapter 5. As stated earlier the quantitative task performance data showed that the groups who had access to all modalities (visual, haptic, audio) solved the test task significantly faster than the ones having access to only visual and haptic feedback. While this is an important result it does not in itself explain anything. When considering the collaboration and especially the dialogues in the visual/haptic and visual/haptic/audio groups a number of differences were found shedding light on the reasons behind the quantitative data. One such difference was the need to ask clarifying questions in the visual/haptic group. The questions posed, both by the sighted and non-sighted participants, for clarification purposes indicated a lack of awareness of the other person’s actions. This in turn can, to a large extent, explain the differences in task completion times between the groups.

The combination of modalities was also altered in the experiment described in chapter 6. In that case it was shown that the task of filling out 10 empty cells in a complex topology of cubes was performed significantly faster in the visual/haptic/audio condition than in the visual/haptic and visual/audio conditions. Thus, this was the only thesis study in which three combinations of modalities were compared against each other and the only one that specifically showed that haptic feedback could make a difference when added to a visual/audio interface. This is another one of the main contributions from the thesis and once again a detailed analysis of behaviors and dialogues gave more insight into why haptic feedback made a difference. The situation tested was quite artificial since it did not directly correspond with any real situation, and thus generalizations of the results to other cases should be made with care. This being said, the study illustrates how audio and haptic feedback may be used, and how these senses complement each other, in a situation where two sighted users are collaborating in an interface with visually occluded target areas. The study also confirms the usefulness of the triangulation used in the earlier studies.

The most important general conclusion drawn from all thesis studies, with regard to modality combinations, is that the combination of modalities influences the collaboration and especially the communication between two users solving joint tasks in the multimodal interfaces developed within the scope of the thesis project. The same conclusions have been shown, or at least indicated, within different application areas.
Even though it is hard to generalize the results gained to other application areas than the ones studied within the thesis scope, the results indicate that there could be a clear gain from studying the same issues within other areas as well.

7.2 Multimodal communication revisited

The most important theme addressed in this thesis is the one concerning the influences of haptic and audio feedback on communication between two users in a collaborative multimodal environment. Several haptic guiding functions and one audio guiding function were tested in the studies presented earlier in this thesis and there are strong indications that the use of these functions affects both the qualitative and quantitative variables analyzed.

7.2.1 Specific functions for communication

As stated in section 4.3.2 the haptic functions used in the study involving visually impaired pupils, as well as in the follow-up study with blindfolded participants, was inspired by functions earlier implemented by Sallnäs (2004). In the earlier study it was shown that haptic feedback could play a major role when used in a function for passing objects to each other in a virtual scene. A function used for grabbing the other user’s proxy was also shown to influence the feeling of being together in the same environment. In the thesis studies these haptic communicative functions were improved (forces were made to be more distinct) and used for new purposes. The joint handling of objects was shown to be beneficial also in the thesis studies, but these studies also showed that there are other ways in which these functions can be used in collaboration.

One could argue that the study involving blindfolded participants gives weaker results when it comes to the usage of haptic feedback – these participants are used to relying on the visual sense (i.e. vision dominates touch) and are probably not used to utilizing their sense of touch for navigation, exploration and communication. When judging the results it is important, though, to bear in mind that the results regarding the use of haptic feedback for communicative purposes were very similar in the study involving visually impaired pupils. In this way the study with visually impaired pupils validated the results of the follow-up study at the same time as the follow-up study generated more qualitative data. Hence, from the first two thesis studies it is clear that haptic feedback can be used for communicative purposes in collaboration between persons who can see and persons who cannot see.

As stated in the literature review in chapter 4.2 haptic functions for guiding have been used before to some extent in situations where sighted and visually impaired children have been collaborating (Rassmus-Gröhn et al., 2007; Magnusson et al., 2007). In their studies the haptic guiding function was initiated by the sighted pupil, controlling a mouse pointer. As long as one of the mouse buttons was pressed, the proxy controlled by the haptic device (by the visually impaired pupil) was bound to the mouse pointer position. This made it possible for the sighted pupil to guide the visually impaired one around the drawn figures and shapes. In this way the functions used in their studies are similar to the
ones used in the thesis studies both when it comes to functionality and general purpose. There is, however, a major difference in that both participants are in control of a haptic feedback device in the studies in this thesis. When initiating a guiding operation by means of a mouse it is impossible to feel response forces from the PHANTOM user. Thus, the guiding function will not be reciprocal in that both users experience the same feedback. If both users do not experience the haptic feedback it is impossible to gesture and more importantly expressing intentions during the guiding process.

The thesis studies illustrate that reciprocity in the guiding functions is a key factor when using haptic feedback for communicative purposes, especially when one of the collaborating partners cannot see. In this regard it is interesting to compare the guiding functions used in the thesis studies to the ones used by Oakley et al. (2001) in their study of haptic feedback in a shared text editor for designers. The application areas cannot be compared, but in this case the particular functions are in focus. See section 2.3.2 for a description of the study and the different haptic communicative functions used. In the study performed by Oakley et al. (2001) the purpose of most of the haptic functions used was to join each other in the same location in the workspace. When one person wanted the other to look at his/her work he/she could push a PHANTOM button to “drag” the other user to his/her proxy position by a constant force. As an alternative a user could locate the other one and move to his/her position in the workspace by pushing a PHANTOM button and thereby being “dragged” by a constant force to the other user’s position. In both these cases the guiding was controlled by the system in that a constant force was used which none of the users could control. Plus, even though both users had haptic devices, the guiding functions were not reciprocal – the force ended when the two proxies were collocated.

A result presented by Oakley et al. (2001) was that participants in some pairs in their study refrained from using these functions mostly due to social protocols – you do not just grab someone and force them to move to your own location. The pairs that did use the function did however gain better results. The guiding functions used in the studies in this thesis did not seem to be problematic in this sense, despite the fact that a force was applied between the users’ proxies in both cases. There is an important point to be made here about the nature of haptic communicative functions. There is a big difference between a constant force guiding (used by Oakley et al. (2001)) and the real time continuous guiding functions used in the thesis studies. If guiding is forced the users have no control over the guiding process, except in the initialization. The one being guided to a particular place has no choice but to follow along. When guiding is experienced continuously and in real time both users feel forces from each other’s proxies all the time during the scope of a guiding operation (regardless whether it concerns moving a common object or directly holding each other’s proxies). One important consequence of continuous guiding is that intentions can be expressed through forces applied to the other user directly or through a commonly held object. If the one being guided does not want to follow along, or wants to go in another direction, he/she can always apply forces in the direction in which he/she wants to move. Communicating intentions directly to the other user in this way, by e.g. resisting movement, is not possible in the case of constant force guiding. The above mentioned differences between constant forced guiding and continuous real-time guiding might explain the differences in acceptance between the
guiding functions used in the study performed by Oakley et al. (2001) and the ones used in the thesis studies. More studies are however needed before this can be proved.

Continuous guiding functions, like the ones used in the thesis studies, enable two-way communication in a way that enables a sighted person and a person who cannot see to collaborate in solving simple construction problems. It is not hard to conclude that solving construction tasks would be much harder if real-time continuous haptic communication could not be used – a constant drag force between two places in the interface would result in both users missing important subtle cues about intentions. In this thesis it has been shown that haptic real-time communication can be used in a straightforward manner in a way that makes both the sighted and non-sighted participant a part of a task solving process involving moving of objects. The clear effect that haptic guiding functions have on communication has not been shown before in similar settings.

In chapter 5 one example of an audio guiding function was also studied. It is hard to draw general conclusions from the results of only one study, but the function used nevertheless shows that audio feedback can be used to give information of one’s own position to a non-sighted user or to request the same information from the sighted user. This localization feature was used successfully in the study and was shown to serve a clear purpose. There were, however, some problems associated with the implementation of this sound, and these will be discussed later on in section 7.3.

### 7.2.2 Influences on the dialogue

One thing in common for all thesis studies conducted is the focus on dialogues and then specifically on how the use of haptic and/or audio feedback influences the dialogue. One of the most important results gained from the thesis studies is that haptic and audio communicative functions can replace verbal guiding. As highlighted in chapters 4-5 giving accurate verbal directions to a visually impaired/blindfolded user is cumbersome and as a consequence the person who cannot see is likely to be at least partly excluded from the work process. Clear indications of this were shown in the field study, within the MICOLE project, preceding the study presented in chapter 4 (Sallnäs et al., 2005). They showed that visually impaired and sighted pupils worked in parallel with their own material, to a large extent. When communicating with the visually impaired pupil giving clear and accurate directions, both regarding directions to move and regarding the shared work material, is a necessity. Providing the accurate feedback verbally during joint task solving can be a tedious and demanding task for a pupil in elementary school. This is one of the reasons why the result that haptic guiding functions can reduce and sometimes even completely replace the need for verbal directions is important.

Even though some problems were identified in the developed applications (see section 7.3 for a more thorough discussion about technical concerns) it is clear that the applications aided in the inclusion of the visually impaired pupil in group work with sighted peers. Several other applications were also developed within the scope of the MICOLE project and these are described thoroughly in section 4.2. Since the entire MICOLE project aimed at supporting collaboration between visually impaired and sighted pupils (and in some cases also teachers), all applications developed were based on collaborative virtual environments and several also included haptic feedback. In this
respect the application area, the technologies used and the particular target group, are all quite unique.

There is one thing that distinguishes the work performed within the scope of the thesis from the other work performed within the MICOLE project – focus on the dialogue. In all thesis studies the dialogues have been the most important unit of analysis and especially in the study involving visually impaired and sighted pupils the content analyses of dialogues gave important insights in how to include a visually impaired pupil in joint task solving. By constructing a joint problem space, with the aid of haptic feedback, several of the problems highlighted in (Sallnäs et al., 2005) were addressed. By focusing on the dialogues, and then especially on how common ground is established through e.g. deictic references, the haptic feedback’s effects on the collaboration was made clear.

The study with visually impaired pupils and the follow-up study with sighted and blindfolded sighted students, both showed similar results when it comes to the effect of haptic feedback on the dialogue – the haptic feedback not only replaces the need to provide cumbersome verbal directions, but it also enables the construction of a shared problem space. The collaborators can then refer to this joint space when discussing the next move and when forming strategies, by e.g. giving meaningful deictic references. In the study involving blindfolded sighted participants the effect of sound feedback on the dialogue was also studied. Although the sounds were mostly meant for conveying awareness information about one’s own and the partner’s actions, it was shown that even the sound could replace the need for verbal directions. To my knowledge, no other study has been found to date in which sound has been used to give deictic information about directions in a task solving setting.

Although no specific functions for communication were used in the study presented in chapter 6 the joint problem space created by the use of haptic and/or audio feedback did also influence the communication between the collaborating partners. Several verbal deictic references accompanied by gestures were made to areas that could not be seen by any of the users and thus once again it was shown that haptic and audio feedback can influence collaboration and especially the communication between two collaborators working in the same interface. As was the case for the studies of collaboration between sighted and non-sighted users presented in chapters 4 and 5, no one has to my knowledge studied the effect of haptic and audio feedback on the dialogue before. Thus, here lies one of the substantial contributions of this thesis.

Winberg and Bowers (2004) and Winberg (2006) did study the effect that sound had on collaboration between sighted and visually impaired users, by focusing on dialogue excerpts. The approach followed by them is similar to the one used in this thesis, but the setting for the studies are quite different. Both Winberg and Bowers (2004) and Winberg (2006) studied how two persons, who were provided with different kinds of feedback, could collaborate in solving joint tasks in a visual/audio interface. Thus, the feedback experienced by the sighted user was totally different from the feedback experienced by the visually impaired one. In the thesis studies both users have always had access to the same information (with exception of the sighted person looking at the screen) and the same means of providing input to the applications. The different approaches followed by the references cited above and the studies presented in this thesis will likely have an effect on the dialogues between the collaborating partners. Even though the analysis
method is very similar this still means that the focus on dialogues in interfaces where the users experience the same feedback is unique for this thesis. This being said, an interesting follow-up on the study presented in chapter 6 would be to give the collaborating partners different kinds of feedback (e.g. one user gets access to the audio functions and the other gets access to the haptic functions). Winberg and Bowers (2004) and Winberg (2006) have shown that collaboration is possible under those circumstances, but no one has ever tested how collaboration would be influenced if the two modalities are haptic and audio.

In each of the thesis studies a content analysis of dialogues has been performed and these analyses have yielded interesting results. Common ground, as defined by Clark and Brennan (1991) has been used to guide most of the analysis, and the use of deixis has been the most important grounding strategy studied. In the studies presented in chapters 4-6 the focus, when it comes to the more detailed analysis of dialogues, has been on means of referencing and other cues displaying the amount of common ground. This being said, it would be interesting to extend the work with an even more detailed analysis of dialogue excerpts by e.g. adopting a conversation analysis approach categorizing all utterances and focusing even more on roles and turn taking. Important insights might then be gained regarding how that very detailed analysis method might contribute to our understanding of haptic and audio feedback and their effect on communication and collaboration. The work by Roque and Traum (2008) has been brought up earlier in the thesis and in their work several categories were defined, which can be used when analyzing dialogues. Even though some of the results presented in this thesis have been linked to some those categories, it would be interesting to apply their taxonomy more systematically.

7.3 Technical considerations

During the scope of the thesis project several multimodal applications have been developed. The reason behind the design and development of new software has always been to get a good test environment. As a result of this, care has been taken not to involve complex object behavior (like deformability) and only a few different textures, friction values and object types have been used. Most attention has been devoted to the main functions studied in the thesis – collaborative functions. In this section the applications developed specifically for the thesis studies will be elaborated on further.

As stated earlier, the application developed for the study in chapter 4 was unique in the sense that it made it possible for sighted and visually impaired pupils to work together in solving construction tasks in a visual and haptic dynamic interface. In this sense the application in itself is a contribution of the thesis work even though it is based on applications earlier evaluated by Sallnäs (2004). The application enables joint object manipulation and also direct contact between the two user’s proxies. These two collaborative functions have been shown to add value to the interaction between visually impaired and sighted pupils and this holds true even in the follow-up study with collaboration between sighted and blindfolded persons. The way in which the haptic feedback was implemented showed the power of communication by means of haptic feedback.
As discussed in section 4.8.3 there were problems with especially the dynamic application illustrating a room with movable blocks. The main idea, regarding use of technology, with the first version of the dynamic application was to investigate how haptic feedback in particular could ease inclusion of a visually impaired pupil in group work. In order to investigate this, a central issue to look at was the establishment of common ground between the members in the collaborating group. The focus was entirely on haptic feedback and the results indeed showed the potential of utilizing that feedback type in this kind of application. The problems experienced in the dynamic application – mostly emanating from the lack of awareness information – were partly due to the focus on haptics. One could argue that it was a big mistake not to consider audio feedback in the first version since the guidelines from e.g. Sjöström (2001) and results from earlier work on software support for visually impaired indicate that audio feedback can play an important role. The lack of audio (or other means of getting awareness information), however, did give important insights into how collaboration and communication is influenced by the lack of awareness information. It brought forward a comparison between a visual/haptic and a visual/haptic/audio version of the interface which highlighted how communication and collaboration changes when the audio modality is added. Thus, despite the fact that the original implementation to some extent hindered the group work process, the results were awarding and resulted in new hypotheses to be tested.

As indicated earlier, the audio implementation presented in chapter 5 was not perfect. Audio was shown to give a significant added value but it was a general opinion among participants that it was hard to hear exactly where a sound came from. Even though most participants reported that the exact positioning was not necessary the positioning issue should be explored in more depth. Wearing ear-phones would probably not be a usable solution, especially if communication between the participants is the main interest. In future research it could, however, be worth to test the application with some kind of surround system to further enhance the depth dimension.

The application developed in chapter 6 has already been thoroughly discussed, especially when it comes to the sound implementation. It is worth pointing out again that the design and programming of this application is a contribution from the thesis work, although it does not include any complex haptic or audio functions for communication. This application was entirely abstract and the intention was not to mimic a chosen real world situation (even though there are indeed many interfaces where target areas are visually occluded). The situation was constructed in order to force the participants to use other modalities than vision. Nevertheless, the application served an important purpose – it could be used to show how communication is influenced by the modality combination when no specific functions for communication are implemented and it gave several important insights into how one can utilize the strength of the respective sense. These facts hold true no matter what opinion one might have of e.g. the implementation of the empty scan sound.

A last important fact to consider here is the quality of the haptic feedback itself. It is a well-known fact to everyone working with haptic interfaces that the haptic equipment is not always stable. One problem which is particularly common is fall through – sometimes the proxy can penetrate surfaces/objects and it can even get stuck inside of objects. Both these problems occurred a few times during the tests with the applications used in the
studies presented in chapters 4 and 5. Every time the problem occurred the participants recovered from it in a few seconds and then continued as usual. The application discussed in chapter 6 did not present any technical problems with regard to the haptic feedback. The main reason for this is probably that this application is simpler in that it did not contain any haptic guiding functions.

7.4 Ongoing research on haptics within the medical domain

One of the application areas showing great promise when haptic feedback is concerned is medicine and that domain was studied as a part of the thesis work. Since the studies performed within the medical domain could be seen as pre-studies and since they did not include a collaborative multimodal interface, the medical studies do not play a central role in this thesis. Nevertheless, these studies will be discussed briefly here, since they generated interesting ideas for future research within the medical domain.

The studies performed within the medical domain aimed at developing a haptic tool for supporting radiologists and surgeons in the treatment planning process. The studies presented in chapters 4-6 generated knowledge on how haptic feedback can influence the communication between two persons working in the same virtual environment. One of the aspects in focus in those studies was the use of deictic references. The work performed within the medical domain applied the knowledge gained from the other thesis studies, with special emphasis on deictic referencing.

The specific setting of interest was the so called multi-disciplinary team meetings, in which medical personnel from different disciplines (surgery, pathology, radiology,...) meet to discuss patient cases, on a weekly bases. The setting has been subject to extensive studies by Kane and Luz (2006; 2009; 2011), Burton et al. (2006), Måseide (2006), Li et al. (2008) and Frykholm (2013), to name a few, and most studies conclude that the weekly meetings have a positive outcome for patients compared to when single physicians make decisions alone. The main meeting artifacts of interest for this thesis are medical images, resulting from MR and CT examinations, projected onto large screens in front of the audience. Figure 23 illustrates the setting.

Figure 23. The meeting setting. The radiologists’ workstation can be seen in the lower left corner. Image taken from Sallnäs et al. (2011).

The interaction with the images is controlled by the radiologists and hence the participating surgeons have no means of interacting with the image data presented.
meetings are time critical mostly due to tight schedules for the medical personnel involved, and the issue with control of the image data affects both the efficiency and the means of communicating (Groth et al., 2009).

Sallnäs et al. (2011) reports on a number of field studies conducted at Karolinska Institutet in Stockholm. In a first round, material from old field studies was considered in order to study the means of referencing image material in today’s medical team meetings. It was found that the participating surgeons often gave verbal deictic references to parts of the images, sometimes accompanied by pointing gestures. Such verbal explanations and directions can sometimes be long and rather cumbersome and most of all hard to understand for novice surgeons. Even though the participants usually establish a definite diagnosis for each patient during the meetings it was clear that the means of reaching this consensus decision could be better supported. This is the main reason why the second round of studies in (Sallnäs et al., 2011), where I took part in the execution and the analysis of the results, concerned the meeting situation and how one could develop tools to ease the discussions about medical image content. Three new field studies were conducted at Karolinska Institutet, during three consecutive 1.5-2 hour long sessions, in which laser pointers were handed out to all participants except the radiologists. The main rationale was to further investigate if and for what purposes the meeting participants needed to point.

It is clear from the results gained that there is a need for pointing during the meetings (Sallnäs et al., 2011). Even though only a few of the surgeons used the laser pointer, the ones who did were usually the surgeons responsible for the patient currently under consideration. These surgeons (except for the patients themselves) are the ones that benefit most from a good amount of common ground during the discussions. Several pointing gestures were identified during the meetings and these were denoted; deictic references to detail, deictic references to area, coordination of information retrieval, navigational guidance and representational gestures.

Deictic references to detail were one of the most common pointing gestures identified. In these cases surgeons used the laser pointers to attract everyone’s attention to a specific detail of interest in one of the radiological images displayed. Such details could be very small areas with a slight shift in gray-scale and it is easy to conclude that it could be a tricky task to explain the position of the point of interest verbally.

Deictic references to areas were made by moving the laser dot along a vessel or using the laser dot to encircle a specific area of interest. This type of pointing gesture was also quite common. It was mainly used by the surgeons when explaining which anatomical parts they referred to in their reasoning. As for deictic references to specific points, the areas indicated with the laser dot could have been explained verbally (as was usually done during the meetings), but with higher time costs.

Coordination of information retrieval was used to request additional information from the radiologist, about an aspect of special interest, by pointing at an area while verbalizing the information requested. Thus, this is a pointing gesture which is by definition accompanied with verbal communication. This type of pointing gesture was used when the surgeons wanted the radiologists to retrieve more information from e.g. another examination than the one currently displayed on the big screens.

Navigational guidance is a pointing gesture which is similar to coordination of information retrieval, in that the gesture is meant to guide the radiologist’s actions. In this
case the pointing gesture is meant for guiding the radiologist on what do to in the image currently shown. This type of gesture is used to follow an unfolding object (e.g. a vein) as the radiologist scrolls in the image stack or zooms in on an object of interest.

A **representational gesture** was used whenever the surgeons felt a need to illustrate a certain operation procedure on the images. This type of gesture was used several times during the three meetings studied. Typically, a surgeon drew lines on the images encircling areas to be cut off.

Apart from illustrating that the laser pointers served a real purpose during the meetings, Sallnäs et al. (2011) also describes a new haptic tool for supporting radiologists and surgeon in reaching consensus about patient diagnoses (see figure 24 below).

The study with laser pointers during medical team meetings clearly showed the potential for being able to point at the radiological images shown on the big screens. The study also showed that surgeons in the audience often choose to point at (and follow) anatomical structures like blood vessels. A haptic prototype, which made it possible to feel anatomical structures like bones and vessels, was developed in order to study how haptic feedback can be an aid when surgeons and radiologists are discussing patient cases.

The haptic application shown in figure 24 was developed by graduate student Jonas Forsslund and evaluated by me. The 3D stereographic model shown in the center of figure 24 is composed by connecting the 2D slices of the CT volume with iso-surfaces. In the left and right corner, respectively, the original 2D stacks are shown. When the user scrolls through the image stack on the left the bottom plane moves upwards as indicated by the red arrow. The same correspondence exists between the right image stack and the

![Figure 24. The application enabling the user to touch a 3D representation generated from 2D slices (shown in the upper right and left corner, respectively).](image)
plane indicated by the blue arrow in figure 24. Apart from seeing the 3D representation, with or without 3D glasses, the components shown in the 3D model are also possible to touch and feel by using a PHANTOM device. In this case the voxels are segmented based on attenuation values in the 2D images – those with values higher than a certain threshold are rendered both visually and haptically and the other voxels are treated as “air”.

The application was evaluated with 10 surgeons in single user settings. The evaluation showed that most participants found it rewarding to be able to see the model through 3D glasses. Above all, the stereographic vision made it easier to see distances between anatomical structures in the image. Most participants did not see a clear use for haptic feedback in the current setting (sitting alone, exploring a 3D model by touch), but they saw a clear potential for this kind of haptic tool in education. Some participating surgeons also reported that they found it easier to judge the distances between a tumor and the surrounding vessels when they had access to the additional haptic feedback.

Although short, the study on haptic feedback support within the medical domain shows that the idea of a feelable 3D model is worth investigating further. The single-user study reported briefly in this thesis should just be seen as a starting point for the next step of research concerning haptic feedback within the medical domain. The application developed should also be seen as a proof of concept testing an initial idea – an idea that hopefully will be incrementally refined during the next few years.

Haptic feedback support has been used in medical simulators and other types of training applications for several years now and many of these show potential for haptic feedback (Heng et al., 2004; Williams et al., 2004a; Williams et al., 2004b; Forsslund, 2008; Larsen et al., 2009; Chellali et al., 2011). The typical study in this case seems to compare the original means of learning procedures and/or motor skills against training in a simulator providing haptic feedback. Most of these simulators have had a pronounced effect on performance in several cases and this is important.

The kind of haptic tool described in this section is different in that it is not primarily intended for training specific procedures, but to enhance communication. In this way it starts a new branch in the study of haptic feedback within the medical domain. As the other applications discussed in chapters 4-6, the medical application is meant to support communication by means of haptic feedback. The study reported in this section do not actually investigate support for communication, but comments from the participants points to an area worth investigating further in future research – education. From the initial study performed it is clear that novice surgeons experience a much harder time interpreting medical image data than more experienced surgeons. Interview results from the study on laser pointers also show that novice surgeons find it hard to follow the discussions during the MDTM conferences, since they are not yet fluent with the language used. Earlier studies in other areas have shown that haptic feedback can influence the means of reasoning about and understanding of complex processes (Bivall Persson et al., 2007; Hamza-Lup and Adams, 2008; Davies et al., 2009; Vines et al., 2009; Chellali et al., 2011). Whether the same holds true for the application discussed in this thesis is too soon to tell, but it would be a shame not to investigate the possibilities that seem to lie here.

The results from earlier thesis studies, on the influence of haptic feedback on communication, to a large extent inspired the work performed within the medical domain. The next step could be to continue on the same theme and execute a collaborative study.
where a senior and novice surgeon sit side by side exploring the model by touch while
talking about the different anatomical structures. With careful attention to language use
and deictic referencing in particular, such a study could yield important insights into how
haptic feedback can be utilized to support the communication and hopefully enhance the
flow of knowledge from the senior surgeon. Another alternative could be a student-
student collaboration where exploration of different anatomical structures by touch could
be compared to the more traditional study of medical image material. Such a setup could
potentially show if access to haptic feedback influences the means of reasoning about and
learning of anatomy.

7.5 Some reflections on generalizability

The studies performed within the scope of this thesis concern different application areas
and the tasks used in the evaluations and experiments have often been quite specific and
artificial. Naturally, this limits the generalizability of the results gained. Most of the work
performed has also been basic research in the sense that the main goal has not been to
develop applications for wide use. The applications have been developed as a means for
answering research questions about communication in multimodal virtual environments.

The studies described in chapters 4-5 both concern collaboration between sighted
persons and persons who cannot see. One of the big differences between those studies is
that the first study involves visually impaired participants and the second study
blindfolded sighted participants. The implications of this have already been discussed and
the fact that the haptic interface was almost the same in the studies makes it easier to
compare the results. But nevertheless using blindfolded participants instead of visually
impaired ones make it hard to generalize the results to collaboration between sighted and
visually impaired persons. It is not necessary to make this generalization, however. The
study described in chapter 5 still generates important results regarding the role of sound
for supporting awareness in collaborative virtual environments. The tasks used are
artificial in that simple generic blocks are moved around in a virtual room, but they were
clearly enough to be able to draw conclusions about how one can use haptic and audio
feedback for communicative purposes in these types of environments. Thus, the
applications used and especially the tasks chosen are quite limited in scope, but the
conclusions that could be drawn, regarding collaboration and communication, are more
generally applicable. When it comes to the study described in chapter 4 it is, of course,
easier to connect the results to the target group of visually impaired persons. The tasks
used in that study are also more focused on tasks corresponding to the target group’s
current focus in the schools visited. Even though no definite claims can be made about
the potential for learning, strong indications in this respect can be made at least regarding
the static application. Joint work in that kind of application, for collaboration between
sighted and visually impaired pupils, should definitely be studied in more depth to see if
the understanding of geometrical concepts is improved when collaborative learning is
applied.

The study described in chapter 6 is based entirely on an artificial application in which
abstract information should be arranged in a grid. Accordingly, it is hard to generalize the
results to any application area, even though there are many cases in which some kind of
organization of large amounts of data is needed. In the particular case studied important results were gained both regarding task performance and communication. Even though generalization is hard in this case strong results were gained which are well worth investigating further in future research. This concerns both the use of haptic and audio feedback in visually demanding interfaces and general design research on methods for scanning and placing based on haptic and audio feedback. One important point worth noting here is also that many of the conclusions drawn from this study could also be drawn from the studies described in chapters 4 and 5.

One fact that also needs to be considered, for all studies involving haptic feedback, is the hardware used. In all studies the PHANTOM device has been used and this of course has its limitations, especially since the users are constrained to one point of interaction. Exploring scenes through a probe has been found to hinder object exploration (Lederman and Klatzky, 2004) and this has to be taken into account when discussing generalizability of results. Since all thesis studies involve exploration and manipulation through a probe it is not possible to generalize the results to the real world or situations where the users have access to more interactions points. This being said, it is also important to bear in mind that the PHANTOM is still the most widely used haptic feedback device in research on haptic interfaces. How the choice of device has affected the results is not clear, but since the device is widely used the results can be related to a wide range of multimodal virtual environments.

This section, and the thesis, will be concluded with a more personal remark. As stated in the introduction to most chapters I have developed the applications myself and I have been responsible for transcribing and analyzing the qualitative data gathered. Thus, it is inevitable that my own involvement in the studies have affected the results to some extent. As discussed in chapter 3, about the methods, this is always a problem when working with qualitative data. The analysis of qualitative data is always subject to some degree of personal interpretation and the same goes for the choice of representative quotes and dialogue excerpts. This problem is difficult to tackle, although several measures have been taken to decrease it. The triangulation applied in all studies is an important aspect to consider here. Whenever it has been relevant, numerical measures (i.e. objective data) have been used to complement the qualitative analysis. The same formal procedure, regarding transcriptions and analysis of observational data, has also been followed in each study and the iteratively developed coding schemes have been discussed with other researchers in the respective projects. Thus, even though I have performed most of the analysis and development on my own, care has been taken to always be objective in the research process.
References


