Assembling the Senses: Towards the Design of Cooperative Interfaces for Visually Impaired Users

Fredrik Winberg [1]
[1] Centre for User-Oriented IT-Design (CID)
Royal Institute of Technology (KTH)
Stockholm S100-44 Sweden
fredrikw@nada.kth.se

John Bowers [1,2]
[2] School of Music
University of East Anglia
Norwich NR4 7TJ UK
jmb78@btinternet.com

ABSTRACT
The needs of blind and visually impaired users are seriously under-investigated in CSCW. We review work on assistive interfaces especially concerning how collaboration between sighted and blind users across different modalities might be supported. To examine commonly expressed design principles, we present a study where blind and sighted persons play a game to which the former has an auditory interface, the latter a visual one. Interaction analyses are presented highlighting features of interface design, talk and gesture which are important to the participants' abilities to collaborate. Informed by these analyses, we reconsider design principles for cooperative interfaces for the blind.

Categories and Subject Descriptors
H.5.1 [**Multimedia Information Systems**]:– *audio input/output, evaluation/methodology*. H.5.2 [**User Interfaces**]:- *auditory (non-speech) feedback, evaluation/methodology, theory and methods, user-centred design*. H.5.3 [**Group and Organization Interfaces**]:- *collaborative computing, computer supported cooperative work, evaluation/ methodology*. K.4.2 [**Social Issues**]:- *assistive technologies for persons with disabilities, handicapped persons/special needs*.

General Terms
Design, Human Factors, Theory.

Keywords
Universal (or disability) access, assistive technologies, auditory input/output, sound in the user-interface, sonification, user interface design, CSCW, qualitative empirical methods, interaction analysis, Conversation Analysis, cross-modal interaction, collaboration.

CSCW'04, November 6–10, 2004, Chicago, Illinois, USA.

1. INTRODUCTION
It is widely believed that graphical user interfaces (GUIs) have constituted a considerable advance over earlier command line interfaces to computer systems. Direct manipulation [18] methods of interaction, which GUIs typically afford, are often held to make interaction with computers more accessible than interaction techniques requiring the mastery of arcane keyword vocabularies or specialised manipulation techniques. Indeed, directly manipulated GUIs are often defended against recent attempts to go beyond that interaction paradigm. Scheiderman [19] argues that direct interaction methods with very high resolution screens are preferable for most of the tasks that software agents have been suggested for, while Bellotti et al. [1] point to the continued relevance of GUI interface design knowledge as researchers concern themselves with ubiquitous sensing systems.

However, as a number of writers have pointed out, little of this holds when one considers the needs of blind and visually impaired users[1] [5,7]. Indeed, ironically, the more widespread are innovations in computer interaction which require visual ability, the more visually impaired users are disenfranchised from leading edge technical developments. This is true whether one considers interfaces to single user applications or the kinds of cooperative applications studied in Computer Supported Cooperative Work (CSCW) which is yet to devote concerted attention to the needs of visually impaired users. Most of the technical innovations in CSCW assume that some screen-based interface will be deployed to enable users to cooperate with co-workers. Indeed, some of the most spectacular innovations in CSCW have specifically exploited advances in graphical technology, for example, the body of work on Collaborative Virtual Environments [e.g. 2]. We

[1] Along with many blind persons and action groups, we do not confine our use of the word 'blind' to people with no residual visual ability. This is still consistent with acknowledging that there are many different kinds and degrees of visual impairment. The general arguments of our current paper are intended to be inclusive: we imagine our main conclusions will be applicable to the technology requirements of many users with visual impairments not just to those with no residual visual ability. Accordingly, in common with most of the literature on assistive interfaces, we tend to use the terms 'visually impaired' and 'blind' freely, though we are more specific when we need to be – e.g. when we describe the visual abilities of the individuals we have worked with.

believe that CSCW should begin to rectify this omission, especially bearing in mind the recent initiatives for 'design for all' and 'universal access' [e.g. 21].

Within Human Computer Interaction (HCI) more broadly, there is notable research concern for building systems and interfaces which can support access for the blind or be assistive in some sense. Through a discussion of some well-known contributions [e.g. 12], our paper seeks to identify some fundamental design orientations in this field – assumptions, in particular, as to how interfaces should be designed to support collaboration between sighted and blind users. To examine these assumptions, we present a study of interaction between sighted and blind users in a collaborative task, where the interface is graphical for the sighted user and involves sonification of the task for the blind user. While we focus on issues surrounding auditory interfaces for the blind, many of our arguments should apply to the development of, for example, tactile interfaces and for other user groups with special needs. Indeed, ultimately we argue that reflection on the design of systems and interfaces for special groups of users should be a matter of universal interest in CSCW and HCI, not only for moral and political reasons, but because how such users engage with technology is generally informative of just what we mean by 'interaction', 'interface', and so on.

2. ASSISTIVE INTERFACES

If you are visually impaired, or for some other reason cannot make use of the graphical information presented on the screen, the readily available options are largely limited to various kinds of 'screen reader' – a device that presents the contents of the screen using speech synthesis or Braille. These (temporally or spatially) linear ways of presenting information have difficulty dealing with spatially patterned and/or concurrent information – matters which GUIs have less trouble with. In HCI research, there have been a number of attempts to improve on this situation by enabling blind persons to access interactive features which screen readers tend not to support. Edwards' early work on the Soundtrack word processor [6] was one of the first attempts at giving blind computer users access to GUI-derived features and peripheral devices. The user interacted with the interface using a regular computer mouse, something that has been investigated further by Pitt and Edwards [15].

Giving access to a complete windows system has been the goal in the Mercator [7,12,13] and GUIB projects [13,14], but in different ways. Mercator presents interface objects in a tree structure through sound, while, in addition to an auditory interface, GUIB uses specialised tactile hardware to retain the interface objects' spatial relationships [13,14]. Assistive interfaces which systematically implement some of the constituents of direct manipulation (e.g. continuous presentation of interface objects) have been investigated at length by Winberg and Hellström [22] and also to some extent by Savidis et al. [17] and Petrie et al. [14]. Amongst other areas, assistive interfaces have been proposed for the web, editing mathematical equations and drawing graphs [e.g. 23].

2.1 Design Principles and Orientations

Many of these approaches to assistive interfaces have been application-specific in their vision. This has not encouraged a fundamental deliberation on the nature of their design orientation. It is only with writers such as Mynatt [12] that ambitious development goals (in her case, giving access to the complete X11 windows system) have forced reflection on and identification of generic design principles for interfaces for the blind. Some of the most important principles are as follows. Non-visual interfaces should:

- provide access to *functionality*, in principle all the things that can be done by a sighted user of a GUI should be possible with an interface for a blind user (this will include giving a sense to the 'iconicity' of the interface and its spatial arrangement)
- enable *manipulation* and *exploration*, if possible the direct manipulation methods of GUIs should be given a parallel implementation
- be *coherent* with visual interfaces to enable *collaboration* between sighted and visually impaired users (this is likely to involve the cross-media translation of significant interface elements to create a common mapping of the elements used in both kinds of interface)

These design principles have considerable face validity and, indeed, can be taken as marking out a research program which many of the more ambitious contributions to the field have followed.

2.2 The Question of Collaboration

We wish to examine the question of collaboration in a little more depth as this is of central concern from a CSCW point of view. The importance of the collaborative requirements of blind users naturally arises out of a consideration of everyday work settings where blind people commonly co-work with non-blind people. Given the importance of collaboration between people of different abilities in work practice, it is curious that this is very rarely studied head-on in the HCI literature on interfaces for the blind. Commonly, instead, an 'in principle' argument is given. It is assumed (or hoped) that, *if* a non-visual interface is successful in enabling information manipulation, exploration and provides access to functionality in ways which are coherent in the above senses, *then* effective collaboration can emerge as a natural benefit. That is, if the same relationships between manipulable interface elements are available to both sighted and blind users, then there are no barriers to effective collaboration – at least not on interface design grounds.

Mynatt and Weber [13, p168] make the very strong statement that, if mutually coherent visual and non-visual interfaces are available, "cooperation is thus ensured". Petrie et al. [14, p429] argue that using a multi-modal display to represent spatial arrangements through tactile and auditory information will "enable successful collaboration" with other users who benefit from standard GUIs. Finally, Savidis et al. [17, p118] remark concerning the design of non-visual interfaces which parallel the display relationships of visual ones: "this will inherently lead to equal computer access opportunities for blind users".

The *a priori* flavour of these arguments is sometimes further underwritten by conceiving of collaboration and communication as being adequately enabled when participants' 'mental models' are near-equivalent [7,13]. On this view of collaboration and communication (as critically dependent on the similarity of mental models), it seems reasonable to evaluate designs in terms of their effectiveness in enabling blind individuals to make analogous judgments and perform equivalent actions as their sighted counterparts (albeit with different technical resources). Such evaluations can be performed using classical individual testing methods.

If interface designs are positively appraised this way, researchers have felt able to assume that effective collaboration would naturally follow. Indeed, most of the studies we have reviewed, alongside making optimistic remarks about collaboration, offer evaluations of their technical innovations along these lines.

However, the relationship between individual cognition and social interaction is a highly contested affair both conceptually and methodologically. Many sociological contributions to the field of CSCW, for example, would insist on the irreducibility of social interaction, cooperation and collaboration, and query understandings of communication based centrally on the relationships between participants' alleged mental contents ([20] is an early example of this tendency in CSCW). From a different perspective, Hutchins [11] not only emphasises the 'emergent properties' of socially distributed cognition, he also notes how the properties of different representational media influence what can be represented and communicated using them. Without taking sides on these arguments, they do give us some scepticism over the claims that collaboration between sighted and blind individuals will be 'ensured' as an 'inherent' benefit of the cross-media translation of interfaces (e.g. a GUI to an auditory display). These can only be issues to investigate empirically.

3. THE STUDY

To inform such matters, we have conducted a study of collaboration between blind and sighted persons in which the participants use auditory and graphical interfaces respectively. To make the study maximally informative of how interfaces in these different modalities can be used collaboratively by individuals with different abilities, we selected a task whose nature could be fully specified both in sound and graphically in a manageable manner. We selected the Towers of Hanoi game as studied and sonified as an example of 'auditory direct manipulation' by Winberg and Hellström [22]. In the Towers of Hanoi, a player has three towers and three or more discs are placed on the left tower in size order, the larger at the bottom. The object of the game is to move the discs from the left to the right tower. A disc can be placed on any of the three towers, just as long as a player does not place a larger disc on top of a smaller. [22] shows that the game can be given a full specification in sound alone such that individual players can complete the problem making no more errors than people do when playing traditional physical versions of the game.

Winberg and Hellström's [22] sonification model is exclusively based on sounds associated with the discs – each disc having its own sound differing in pitch and timbre from the others. In order to distinguish which tower a disc is located on, both stereo panning and amplitude envelopes are used. A disc's vertical position is represented by the length of the sound, the higher the disc is placed the shorter the sound. The user interacts with the interface using a regular computer mouse and a pair of headphones. The mouse uses a focus feature in order to track the cursor. The discs on the tower pointed to by the cursor will be louder than the others [cf. 15]. By doing this, the user can focus on a subset of the auditory display without losing track of the rest. There are also short transition tones that tell the user when the cursor is moved from one tower to another. The discs are heard in a rapid sequence with respect to size in an ascending order. This means that no matter where the discs or the cursor are located, the sequence involves all discs being heard. As the sequence is repeated, the whole display is continuously represented. Each repeat of the sequence takes between about 1s and 1.5s to be heard depending on the number of discs in the game. When the mouse is depressed over a tower, the top disc is selected for movement. With the button held down, the disc can be dragged to the target tower and then released. If this is a legal move, the auditory display is updated, otherwise an error tone is heard and the disc returned to its location. See [22] for further details.

It should be clear that [22]'s sonification of the game is fully *functional* in that there is no logical game feature without a sonic counterpart. The sonification enables *manipulation* of the discs and *exploration* of the state of the game. In principle, this representation in sound should *cohere* with a graphical or physical representation and allow *collaboration* on the game between blind and sighted players. Indeed, [22]'s Towers of Hanoi was intentionally designed to manifest these design principles as listed in our introduction. As such, while the game has some artificiality as a task, it is a motivated choice for study of cross-modal, cross-ability collaboration issues in microcosm.

Three pairs of participants (one blind and one sighted subject for each pair) contributed to the current study. Our blind participants were all congenitally blind with no residual vision. All participants had prior knowledge of the Towers of Hanoi game. The two participants sat at a small table at a 90° angle, the blind participant with a pair of headphones and the sighted participant with a small computer screen. The sighted participant could not hear what the blind participant heard, and naturally the blind participant could not see what the sighted participant saw, so there was no overlap in presentation modalities. The participants had one mouse each. The mice, though, shared a common single on-screen cursor. Each session was videotaped. Also included in the recorded video was a small screen shot of the graphical representation in order to show the status of the game at all times. See Figure 1 for a frame captured from one of the videotaped sessions.

Figure 1. Set-up for the collaborative Towers of Hanoi with screenshot of the graphical version inserted bottom right.

The participants were separated in the beginning of the session, in order for them to be introduced to the Towers of

Hanoi using their specific version of the game. After this, the participants were brought together and informed that they were supposed to solve the game together by taking turns in moving the discs. They were warned that there was only one cursor and two mice and that, if they were to move mice simultaneously, the result could be confusing. The participants were instructed to agree moves in the game but to refrain from moving a disc when it was not their turn. The participants played the game first using three discs, then four and finally five discs.

4. ANALYSIS

As our study has a small participant number (in common with many in the literature on assistive interfaces), we do not adopt an analytic strategy which is geared towards capturing statistical generalities and testing precise hypotheses in a quantitative fashion. Rather, we prefer to offer detailed *qualitative* descriptions of the data. In this we follow the orientation of Conversation Analysis [e.g. 16] which has come to be a more or less established methodological contributor to CSCW since the work of Suchman [20] and through the video analytic work of Heath and colleagues [e.g. 8]. Rather than give summary statistics, this tradition prefers to analyse in depth a number of selected transcribed instances which exemplify phenomena of relevant interest. While Conversation Analysis inspired or derived methods are typical in analysis of naturally occurring interaction, the CSCW literature also contains precedents for their use in analysing interaction with prototype technology in a set task [e.g. 4] – as is the case here.

We introduce our transcript conventions as we discuss the examples. However, one convention needs to be pointed out in advance. We give a textual description of actions at the interface in italics. When these actions are extended and concurrent with talk or a pause in a participant's talk, we align the italicised description underneath the transcribed talk or timed pause. The beginnings and endings of such concurrent activities are explicitly noted with pairs of open and closed square brackets respectively. These are aligned to show the moments in, for example, someone's talk where the beginning and ending of a disc move occurred. Interface actions which are more punctate and are not concurrent with talk or a timed pause are notated in angled brackets.

We analyse a series of data excerpts under these headings:

- turn taking
- listening while moving
- monitoring the other's move
- turn taking problems and repair
- re-orientation and re-establishing sense
- engagement, memory and talk
- disengagement.

In each case, we are identifying a topic of core relevance to cross-modal, cross-ability collaboration on the task.

4.1 Turn Taking

Transcript 1 shows two game-turns being made with the talk and bodily activity of both participants being smoothly and rhythmically interleaved.

Transcript 1 (9:35:38 pair 3, 5 discs)

```
B: [(1.8)                                  ]
   [picks up smallest disc from middle,
                disc to left, drops disc]
S: (2.0) yes and [then two to     [three
                 [mouse to middle
B:                                [to one
   and then two to three yes
S: [yes                      [two to three   ]
   [picks up 2nd smallest disc [disc to right,
                                     drops disc]
   mmm and [then
B:         [then I have to take one to three
```

The blind participant (B) starts by moving the smallest disc to the left tower and dropping it. After a two second pause `(2.0)`, the sighted game player (S) moves the mouse to the middle and proposes moving a disc from there. Note how he refers to the middle as `two` and his suggested destination, the right, as `three`. This is a local convention made by these two players so that they have a common way of referring to the three towers in the game. Simultaneously, B latches onto S's utterance by confirming his last move `(to one)` and then, having heard S complete his proposal, agrees with him. S then continues by making the agreed move saying it aloud as he does so (`yes two to three`). S marks the completion of his turn and invites the transition to B's next with `mmm and then`. B responds with a proposed following move in overlap with and continuance of S's talk `(then I have to take one to three)`.

In this example, two game turns are smoothly accomplished. It is notable how the participants interleave their talk and movement, and do so in ways which make it clear to the other both what they are doing and what their ongoing sense of the game is. They talk and reason aloud, saying what they are doing as they are doing it. In this way, the actions of one party come to be seen as consequential for the next (`and then... and then...`). In sequences like this, B's use of the sonic interface is smoothly folded into the making of moves, the talk, the playing of the game. B can hear the changes in the interface occurring as the initial move in the transcript is made, that is, during the 1.8 seconds it takes to concurrently make the move. The continuous sonic presentation for B contributes to smooth game play with both parties delivering timely moves.

4.2 Listening While Moving

In Transcript 1, B had the opportunity to listen to changes in the sonic display while making a move. In the following, we have a clear example of B attending very closely when designing and executing his move. B, having picked up a disc, hovers over an incorrect location for a while, turning his head as if to listen closer.

Transcript 2 (9:33:11 pair 3, 5 discs)

```
B: [there are two] left on one, [aren't there?
   [mouse to left]
S:                              [yes
B: (0.8) puff this will take [time
S:                           [one to two
B: [one                       [to two       ]
   [picks up 2nd largest disc [disc to middle]
   only [possibility
        [disc to right
   (0.4) [(1.0)                             ]
         [turns his head as if listening closer]
S: that [was three, that's two
B:      [disc to middle
   (0.8) that <drops disc> (0.6) is two
```

The transcript begins with B moving the mouse to the left, and at the same time confirming with S how many are placed there (`there are two left on one, aren't there?`). Again, remember that these participants refer to the towers as one, two and three, corresponding to the left, middle and right tower. S confirms this and suggests a move (`one to two`). B picks up the second largest disc from the left tower (one), and moves it to the middle (two), while repeating what S said (`one to two`). B continues to move the disc, and takes it to the right tower, still holding the disc. B then turns his head, as if listening closer, for one second, and then returns the disc to the middle at the same time as S notes this (`that was three, that's two`). B now drops the disc on the middle tower.

The sonic display enables B to realise when he is about to overshoot his gesture. Note also how B makes his careful scrutiny of the display a public affair. His head turning is noticed by S and this, together with his pause over the incorrect, non-agreed destination, prompts S to point out `that was three, that's two` just as B pulls back to complete the gesture which comprises the agreed move.

4.3 Monitoring the Other's Move

Just as the interface enables blind participants to shape their own moves, we have examples of moves made by sighted players being monitored through their sonification. Transcript 3 shows an example where B monitors a move made by S and displays her realisation that the move has been done immediately upon its completion.

Transcript 3 (17:27:05 pair 2, 3 discs)

```
B: [(2.6)                        ] here
   [picks up smallest disc,
    disc to right, drops disc]
S: mmm (1.0) then [I'll take the left and put]
                  [mouse to left             ]
   it down the middle (0.4) <picks up middle
   disc>
   [(2.0)                        ]
   [disc to middle, drops disc]
B: yes you have it there, yes
```

The transcript begins with B moving the smallest disc to the right, ending the move by confirming this to S (`here`). S then initiates his move by telling B what he is doing (`then I'll take the left and put it down the middle`), while moving the mouse to the left. S picks up the middle disc, and moves the disc to the middle tower and drops it there. Immediately following this move, which takes two seconds, B responds by saying `yes you have it there, yes`. Again a pair of game turns is smoothly accomplished, one made by each participant, the sonic interface enabling B to hear moves made by S as they are made.

4.4 Turn Taking Problems and Repair

From our examples so far, it appears that the sonic interface enables blind participants in the study to take turns with their sighted counterparts in smoothly exchanging game moves. However, this is not always the case. In our corpus, we have a number of examples of a participant prematurely projecting the closure of a game turn made by the other and initiating the beginnings of the next move or reflection upon it by making preliminary mouse movements. This can be problematic because the onscreen mouse pointer is shared and the interrupted move may be completed erroneously. In Transcript 4, the blind person (B) initiates elements of a new move after a pause in audible activity by the sighted person (S) during which a disc is being held over its destination though not yet released.

Transcript 4 (17:35:26 pair 2, 5 discs)

```
S: <picks up smallest disc from right,
    disc to middle>
   (2.0)
B: <mouse to right>
S: <drops disc>
B: but it should be, no <mouse to middle>
S: (0.4) [mouse to right
B:       [I see, you hadn't [moved
S:                          [picks up smallest
                             disc
B: [yet,sorry
S:                     [there-you-go
   [disc to middle  [drops disc
```

At the beginning of the transcript, S starts to make a move by depressing the mouse button and moving the onscreen disc to the middle of the display, holding the disc there for two seconds. However, before the move is completed, B moves her mouse to the right. S then releases his mouse button only to leave the disc on the right hand tower not the middle as intended. Immediately, B notes that something has gone wrong (`but it should be`) and returns the mouse to the middle of the display. After a short pause, S begins to correct the situation by moving the mouse to the right and picking up the rogue disc while B shows that she has realised that she broke into S's move and apologises (`I see, you hadn't moved yet, sorry`). S announces when he has completed the repair (`there-you-go`).

In this example, it is worth repeating that B *immediately* notices that the game has not gone as intended. This must be on the basis of a heard change in the sonic display which does not correspond to the gesture she has just (prematurely) initiated. In other words, the sonic interface enables B to detect the anomaly and make it plain to S that she is aware that his move was not completed. While the sonification of the game does give B resources to detect this anomaly in game play, it is clear that the error has its origins in other perhaps less felicitous features of design (e.g. the sharing of the mouse cursor between two devices) – a matter we will return to later.

4.5 Re-Orientating & Re-Establishing Sense

At the beginning of a turn in the game to be executed by a blind person, we commonly observe the sonic interface being used to re-orient the blind player in the game and re-establish the locations of the discs and the whereabouts of the mouse. In Transcript 5, after S completes a move, B utilises the sonic interface first to locate the mouse cursor, then to pick out the disc that should be moved next.

Transcript 5 (17:31:07 pair 2, 4 discs)

```
S: no ok, [I'll put it on            ] the
   middle
          [picks up smallest disc,
             disc to middle, drop disc]
   instead like that
B: yes it's good that you do as I [say,
                                  [mouse to
   right
   I think that's good (1.8)
   [where am I?     [there is one (3.4)
   [mouse to middle [mouse to right
   let's see if I do as I say myself (2.4)
   [(3.0)                      ]
   [picks up 2nd smallest disc,
        disc to left, drops disc]
```

Transcript 5 starts with S talking through a move and indicating when it is finished (`no ok, I'll put it on the middle instead like that`). B then humorously remarks that `it's good that you do as I say` before moving the mouse to the right. She then seems to be lost (where am I?), and moves the mouse first to the middle and then back to the right again. During the move back to the right, B locates a disc (there is one) and listens to the display for several seconds during which time she rejoins her earlier humorous talk with let's see if I do as I say myself. She then moves the disc she has located to the left. In this example, it takes B a little while to re-establish the state of the game and her position within it (where am I?). However, the sonic interface is an adequate resource for helping her to do so. All the time, though, she is talking aloud while browsing it, giving S opportunities (here withheld) to correct any mistaken reasoning.

Transcript 6 manifests a similar interweaving of talk and browsing the display to re-engage with the game at the beginning of a blind person's turn. At the outset of the excerpt, B asks S whether he has completed his move and checks the sonic interface to confirm that this is the case. She then systematically investigates the display, saying what she finds where, before preparing and making a move.

Transcript 6 (17:35:35 pair 2, 5 discs)

```
B: have you moved it now?
S: [yes
B: [you must [have because this is] empty,
             [mouse to right      ]
   isn't it?
S: yes
B: [(1.0)           ] here are those two?
   [mouse to middle]
S: mm
B: (1.0) [yes they      ] are because there I
   hear
         [mouse to left]
   the third now (1.0) yes the [middle size now
I
                                [picks up the
                                 middle disc
   [move it farthest away then]
   [disc to right             ]
   (0.4) <drops disc>
S: yes
```

B re-establishes her sense of the game's current position having confirmed that S has completed his move. As B moves the mouse across the display – first to the right, next to the middle, finally to the left – she offers her understandings formulated as questions (this is empty, isn't it? here are those two?) which S makes minimal affirmative responses to (yes & mm). As she reaches the left tower, what she hears there confirms her reasoning that there are two discs placed on the middle tower (yes they are because there I hear the third now). This enables her to decide on a move as she earlier confirmed that the right side is empty and hence must be the destination for a disc moved from the left (yes the middle size now I move it farthest away then). This deduction is supported by two notable periods of close listening to the display each for about a second. After both of these B explicitly confirms her emerging understanding of the state of the game (yes they are & yes the middle size) – the first time making it explicit this is based on what she hears (because there I hear the third now).

4.6 Engagement, Memory and Talk

We have argued that the sonic interface is prominently used by blind participants to re-establish their sense of the game and re-orient themselves within the display at the beginning of their game turns. This is done through taking up and moving the mouse across the display while locating and listening at the relevant places. A prominent feature of many of our examples is how participants talk aloud to each other, not merely when discussing what game move to make, but also while establishing and checking their understandings of the state of the game. We have seen that interaction with the sonic interface by blind participants is often accompanied by talk which, potentially, the sighted person could correct or elaborate on.

Transcript 7 examines the reciprocal situation where it is the sighted person's move. Instead of investigating the display through mouse movements, we see B discussing the state of the game with S from memory and on the basis of the sounds of S's movements, matters which becomes problematic at a few junctures.

Transcript 7 (17:35:49 pair 2, 5 discs)

```
B: <removes hand from mouse>
S: (2.0) but the question is what you (0.8)
B: the small is in the middle, isn't it?
S: yes
B: it must go to far [right, doesn't it?
S:                   [picks up smallest disc
   (1.0) <disc to right, drops disc>
B: (0.6) or am I getting lost now?
S: yes, [no::             [n::yea::h
        [mouse to middle
B:                        [yes but I did, it
  should not, it should go to far left
S: (2.0) <mouse to right>
B: you do have two in the middle now?
S: no, there is one in the middle
B: (1.4) yes, but the small, it should go to far
   left, shouldn't it?
(3.0)
B: because it is (0.6) yes
S: [yes,                    ok [now I get it,]
   [picks up smallest disc    [disc to left ]
   that's correct
B: like that?
S: <drops the disc> like that, yes and [then
B:                                     [yes
   it has to be like that
S: then the middle, the
   [second largest should go to the right then
B: [mouse to middle
B: yes, it's here, isn't it?
S: yes
B: (3.0) <picks up 2nd smallest disc, disc to
         right, drops disc>
```

The transcript begins with an incomplete utterance from S (but the question is what you) followed by a pause indicating that he is not sure what to do. B then begins to reason about the move to be made by first confirming the location of the smallest disc (the small is in the middle, isn't it?), and then entertaining the possibility of moving of it (it must go to far right, doesn't it?). S takes this as a suggestion for a move and makes it, initiating his move of the disc during B's talk. However, B now is unsure over her reasoning for moving the disc to the right (or am I getting lost now?), an uncertainty shared with S (yes, no:: n::yea::h). B corrects herself (it should not, it should go to far left) and after a two second silence queries the state of the game (you do have two in the middle now?). There *were* two discs in the middle before S made the last move, but now there is just one, as S points out (no, there is one in the middle). B then reaffirms the suggested move (yes, but the small, it should go to far left) in a formulation completed with a tag-question (shouldn't it?). S withholds reply, so B hesitantly begins an account of why she is making this suggestion (because it is (0.6) yes). S follows by moving as suggested while saying yes, ok now I get it, that's correct. B and S then confirm with each other that the correct move has been made and that they both understand that this is the case (B: like that? S: like that, yes… B: yes it has to be like that). This contrasts with the earlier incorrect move of the small disc when S made the move after a more uncertain suggestion from B and without either party confirming that the move had been made or checking that the other had realised it. Having re-established a mutually understood sense of the game, S says what the next move should be (then the middle, the second largest should go to the right then) during which B moves the mouse to the middle as S suggests. She then confirms that she is in the correct location to start making the move (it's here isn't it? S: yes) and after three seconds of listening to the display transfers the disc to the right tower.

For most of this excerpt, it is S's game turn. He moves once, then corrects it. Throughout, B is partaking in the discussion of the move to play but worries that she is getting lost and at least once is mistaken about the state of the game. This is not helped by S making a move without clear confirmation from B that it should be made. Furthermore, when S makes the first move he does not talk his way through it. Through this time, S alone manipulates the mouse. B does not interactively investigate the game through browsing the sonic display but rather relies on her memory of the game and S's answers to questions. By contrast, when B takes up the mouse again in her next turn, she is able to quickly locate the desired disc and fluently make a move. Transcript 7, and the contrasts it makes with some of our earlier examples, suggests the importance of ongoing interactive engagement with the sonic display. It is relatively easy for the blind participants to work out the state of the game when they are able to investigate its contents through moving the mouse. Differences between the different locations will be readily heard through comparison. These can be further checked through talking aloud or directly asking the sighted co-player. However, when it is the sighted person's move, even though the state of the game is sonified in the continuously presented auditory display, the display is only available for manipulation on pain of confusion over mouse control and pointer locus – a matter which can disrupt collaboration (see Transcript 4). Accordingly, the blind person's reasoning about the game becomes more resourced by memory and talk than active display manipulation.

4.7 Disengagement

Our analysis of Transcript 7 would suggest that, if a blind participant were to spend some time disengaged from the active manipulation of the display, then their understanding of the state of the game and their reasoning about it might become notably impaired. Transcript 8 is a little longer than some of our other examples but is relevant to this issue. During the course of it four game-turns are made – the first *three* by the sighted person, S. This departure from the game's turn taking is initiated by the blind particpant, B, following some confusion at the end of S's first move. Notably, B's disengagement from the active use of the display makes it problematic for him to re-acquire a sense of the game and re-establish his ability to take turns in it as expected. This is the case even though S carefully describes the location of the discs to B throughout.

Transcript 8 (9:34:04, pair 3, 5 discs)

```
S: [three to one                                  ]
   [picks up 2nd smallest disc, [disc to left,
                                       drops disc ]
B:                              [mouse to left    ]
S: now both of us move at the same time <laugh>
B: no?
(1.6)
S: now we have (0.4) we move three to one,
   now we take two to one
B: yes, right
S: you or me?
B: you
S: my turn [(2.0)                           there]
           [mouse to middle, picks up smallest
                  disc, disc to left, drops disc]
   two to one
B: what do you say, what have we now? now we
   have on [(0.8)           ]
           [mouse to middle]
S: on peg one (0.8) we have the (0.4) largest
   (0.4) and then the (0.4) yes the two smallest
   on it [then
B:       [the two smallest yes
S: and [on the middle peg] we have (0.4) the
B:     [peg number two    ]
S: second largest and to right we have, well the
   third largest, the middle so to say, so it
   will be three to two, have to be
B: yes it have to be
S: yes
B: mmm
(3.0)
S: <mouse to right>
B: it is you
S: [three               [to                [two
   [picks up middle disc [mouse to middle [drops
                                            disc
B: then number three is empty now?
S: yes. [then-we'll
B:      [now let's see one must over to (0.4)
S: three
B: mmm, right [(1.0)         ][then we take it
              [mouse to left][picks up smallest
                                 disc
   [and              ] what did we say? to?
   [mouse to middle]
S: to three
B: [to three yes (1.5)] [we put it down here
   [mouse to right    ] [drops disc
```

As S completes his move of three to one at the beginning of the transcript, B moves his mouse to the left. Because this is the same direction as S is moving, B's simultaneous use of the mouse doesn't interfere with the mouse-focus in the manner shown in Transcript 4. The auditory display would not highlight anything anomalous about this concurrent use. However, S sees B move the mouse on the table top and is able to remark now both of us move at the same time and laugh. Given what he can hear, this observation doesn't have a clear sense for B (no?). After a 1.6 second silence, S hesitantly states the game move just made (now we have (0.4) we move three to one) and suggests the next (now we take two to one), to which B agrees. S queries whose turn it is and B (incorrectly) tells S to take it. S makes the move and, only once it is complete, says what has been done: there two to one.

B now tries to re-establish his sense of the game (what have we now? now we have on) and begins to explore the display with a mouse move to the middle. S breaks into a pause in B's talk, while B is moving the mouse, and begins a full description of the display. B stops exploring the display through moving the mouse and twice echoes terms in S's description (the two smallest… peg number two). In neither case does B draw any implication on the basis of S's emerging description, nor does he formulate a proposed move. Indeed, it is S who reasons what the next move should be (`so it will be three to two, have to be`). B agrees. However, the actual execution of the move only follows after a long silence which S breaks with a hesitant mouse movement. B again tells S to take the game move which S this time completes in a style which latches the move's components to components of his talk. Picking up the disc is accompanied by `three`, movement of it occurs while uttering `to`, and `two` is said as the disc is dropped.

Immediately following this move, B draws a correct implication: `then number three is empty now?` S agrees and begins to suggest a next game move (`then-we'll`) in overlap with B who continues by identifying `one` as the disc to move. After a short pause in B's talk, S completes the specification of the move of disc one to `three`. B picks up the disc from the left, pauses in the middle and checks its destination with S (`what did we say? to?`) before finally moving to the right (`three`) and releasing it.

In this example B passes over the opportunity to make a move on at least two occasions. This includes one case where the move is clearly agreed between the participants and just needs to be made – B preferring to let S do it. However, this should be understood not in terms of the sighted participant dominating the game or not showing enough care in helping the blind co-participant understand how things are. The transcript contains many instances where S holds back in making a move until prompted to do so by B. Equally, S offers detailed descriptions of the state of the game. Rather, it is the link between display manipulation, auditory display and the game which has been broken for B – quite possibly as a cumulative result of the confusion at the beginning of the excerpt. Furthermore, the longer it remains S's (unofficial) turn, the more B's re-engagement with the display is deferred and the more difficult that, in turn, may become.

5. DISCUSSION

We have presented a series of qualitative analyses of pairs of persons, one blind, one sighted, collaborating in a game on a move by move basis. Our blind participants have the game represented to them by a sonic interface, their sighted counterparts having a corresponding GUI. In microcosm, this setting has allowed us to examine some basic issues to do with cooperation between people of different physical abilities supported by interfaces in different modalities. Our analyses suggest that sonic interfaces can be designed to enable blind participants to collaborate on the shared game: all pairs completed all games[2]. While this testifies to Winberg and Hellström's [22] sonification design, more important for the concerns of this paper is capturing *how* the sonic interface is used as revealed through our qualitative analyses. Here we connect with some general issues and go beyond the specifics of the study. Many of these are familiar from the CSCW literature but have not before been connected to questions of technology accessibility.

We have seen that the sonic interface can help blind participants smoothly interleave their talk and gestures at the interface (browsing, moving and so forth). The sonification of the game, as it follows a principle of continuous presentation of interface elements, enables the conduct of co-players to be monitored. It also supports the design of blind players' own gestures (e.g. a game move is *heard* to change the display, and hence the state of the game, *as the move is made*). In these respects, the interface helps the blind person participate in a *'working division of labour'* [cf. 10]: while each participant at a particular moment has a job to do (as given by the turn taking protocol of the game), they have resources to monitor each other's conduct and help each other out if required.

However, importantly, the sonic interface is not the blind person's *only* resource for participation. The sounds that it makes occur *as gestures are made* exploring the interface and *as talk is exchanged* to reason aloud or describe the state of things. That is, blind persons interweave their active manipulation at the interface with talk and game play. When they disengage from manipulating the display and listening to the consequent changes, the state of things may become opaque – with memory and careful descriptions from a co-participant barely enough to recover affairs. This is exaggerated by the turn taking regime of the game and that there is only one mouse cursor. Blind participants have to re-establish their understanding of the game and their orientation in it at the beginning of game turns. If a transition between participants has been inelegantly dealt with (e.g. if both mice are momentarily moved), it may be problematic to recover a sense of the game – even leading to an extended disengagement from taking a turn in it.

5.1 Rethinking the Design of Assistive Interfaces

These remarks help us cast the principles for the design of assistive interfaces that we identified from the literature at the beginning of our paper in a new light. Indeed, we can make initial suggestions of ways to go beyond them informed by a direct concern for supporting cooperation.

Manipulation and Exploration. The manipulability of an assistive interface is an essential matter, not just to get things done, but to enable an integrated cross-modality sense of the state of things to be established. It is through concerted manipulation, listening and talk that the display comes to be understood as, indeed, an effective sonification. This effectiveness can be disrupted if the *linkage* between gesture, sound and the state of things becomes unreliable. Maintaining such linkages then becomes criterial for assessing design alternatives. In the current case, sharing the interaction focus across two input devices disrupted linkages and interfered with 'turn design' in ways which would not have been so acute if both participants were sighted. Of course, shared-cursor problems are well known in CSCW and not unique to our study. However, it is important to realise how difficult such problems can be to repair when users have different abilities as *the problems themselves may be identified and experienced differently*. For our blind users, cursor-sharing sometimes disrupted not just the fluency of turn-taking but the delicate linkages between gesture, sound, talk and game-play needed to maintain an understanding of the state of things. Amongst other matters, phenomena like these suggest that issues in the

[2] While our methods have been qualitative, a short quantitative remark is perhaps useful here. The game has a single optimal method of solution in each of its versions. This enables the computation of an error rate (the number of moves which have to be taken back). In terms of this measure, where cross-study comparisons can be made, our collaborating pairs are performing the game as well if not better than [22]'s individuals did. More detailed quantitative analyses are the topic of another study.

design of groupware such as concurrency control [9], need to be fundamentally reappraised if the group has mixed physical abilities and/or presentation modalities.

Functionality. The goal of reproducing the entire functionality of, say, a GUI should be seen as relative to the importance of supporting blind people in combining resources and skills (talk, gesture, listening, remembering) to gain an understanding of the state of things and cooperate with others. One can certainly imagine cases where exhaustive functionality, no matter how well motivated on *a priori* or moral grounds, might obstruct this. Indeed, such considerations might provide a limit case on the viability of some approaches to sonification: for example, if an auditory display takes so long to listen to that it requires too long a time out from talk with a co-participant.

Coherence and Collaboration. Earlier we saw various authors arguing for the support of collaboration through assistive interfaces which cohere with those used by sighted individuals. It is commonly argued that interfaces which have parallel relations between elements in them, albeit in different media, will enable users to develop similar 'mental models'. Some authors claim that this is sufficient for effective collaboration. The sonic and graphical interfaces we have studied do indeed manifest coherent cross-media relations. However, this *in and of itself* does not guarantee successful collaboration. It is essential to investigate *how interfaces are used* and *how such uses are folded in with the varied things co-participants do* (designing gestures, monitoring each other, establishing the state of things and one's orientation in it, reasoning and describing). An assistive interface should be seen as a *resource* for, in and alongside, those activities. It is not enough to merely evaluate the information it provides for forming a mental model. Studying *just how* people in practice use an assistive interface in collaborative settings becomes an important empirical topic for CSCW, no matter how much, in principle, the interface is based on traditional formats or passes individual usability tests.

5.2 Cooperative Interfaces and Assembling the Senses

From time to time, CSCW research has pointed out the cooperative, social interactional dimensions of interfaces even when those interfaces are designed individual use. [3] and [8] show persons (respectively: patients or callers to a telephone banking centre) being attentive to how others in their presence engage with technology (doctors and medical systems or call-takers and bank mainframes). Accordingly, [3] and [8] argue for the re-appraisal of traditional HCI issues from a CSCW perspective – for example, by extending our notion of what counts as a 'cooperative interface'. Further, CSCW research has commonly urged recognising all the activities of situated actors and how these are interleaved on a moment by moment basis thereby exceeding narrow definitions of users' 'tasks' [3,8,10,20].

We have extended these emphases into a preliminary consideration of collaborative settings where individuals with mixed abilities work with technologies which engage different sensory modalities. The conduct of blind and sighted persons in collaboration has been understood in terms of the varied and variably resourced activities they engage in. This essentially involves 'assembling the senses' as what can be heard, touched, moved around, and (if one is able) seen are combined, both through manipulating interfaces and social interaction, to give sense to the world: *interaction as assembling the senses*. We have begun our exploration with a study of collaboration in a 'micro world'. Even here we have found persons to be interacting with interfaces and each other in ways which made us re-think 'accessibility' in interface design from a CSCW standpoint: *assistive interfaces as cooperative interfaces*. We suggest that, by adding detail to our two italicised 'slogans', CSCW research could make a timely and characteristic contribution not just to 'universal access' but to our basic understanding of what we mean by such foundational terms as 'interaction' and 'interface'.

6. ACKNOWLEDGMENTS

We thank all the volunteers who participated in our study. We are especially grateful to Sten-Olof Hellström for his work on the sonification model for the Towers of Hanoi used in this study. Finally, we thank Tony Stockman and five anonymous reviewers whose feedback has helped improved this paper.

7. REFERENCES

[1] Bellotti, V., Back, M., Edwards, W. K., Grinter, R., Henderson, A. and Lopes, C. Ubiquity: Making sense of sensing systems. *Proceedings of CHI 02*. ACM Press, NY, 2002.

[2] Benford, S., Bowers, J., Fahlén, L., Mariani, J. & Rodden, T. Supporting cooperative work in virtual environments. *The Computer Journal, 37*, 653-688, 1994.

[3] Bowers, J. & Martin, D. Machinery in the new factories. *Proceedings of CSCW 2000*. ACM Press, NY, 2000.

[4] Bowers, J. & Pycock, J. Talking through design. *Proceedings of CHI 94*. ACM Press, NY, 1994.

[5] Boyd. L. H., Boyd, W. L., & Vanderheiden, G. C. The graphical user interface: Crisis, danger and opportunity. *Journal of Visual Impairment and Blindness*, (1990, December), 496-502.

[6] Edwards, A. D. N. Soundtrack: An auditory interface for blind users. *Human-Computer Interaction*, 4, 45-66. 1989.

[7] Edwards, K., Mynatt, E., & Stockton, K. Access to graphical interfaces for blind users. *interactions, 2*, 1 (Jan. 1995), 54-67.

[8] Greatbatch, D., Heath, C., Luff, P. & Campion, D. Interpersonal communication and human computer interaction. *Interacting with Computers*, 5, 193-216, 1993.

[9] Greenberg, S. & Marwood, D. Real time groupware as a distributed system: concurrency control and its effect on the interface. *Proceedings of CSCW 94*. ACM Press, NY, 1994.

[10] Hughes, J., Randall, D. & Shapiro, D. Faltering from ethnography to design. *Proceedings of CSCW 92*. ACM Press, NY, 1992.

[11] Hutchins, E. *Cognition in the wild*. MIT Press, 1996.

[12] Mynatt, E. Transforming graphical interfaces into auditory interfaces for blind users. *Human-Computer Interaction*, 12, 7-45, 1997.

[13] Mynatt, E. & Weber, G. Nonvisual presentation of graphical user interfaces. *Proceedings of CHI '94*. ACM Press, New York, NY, 1994.

[14] Petrie, H., Morley, S. & Weber, G. Tactile-Based Direct Manipulation in GUIs for Blind Users. *Conference companion to CHI '95*. ACM Press, New York, NY, USA, 428-429, 1995.

[15] Pitt, I. & Edwards, A. Pointing in an Auditory Interface for Blind Users. *Proc. IEEE Conference on Systems, Man and Cyvernetics*. IEEE, NY, 1995.

[16] Sacks, H., Schegloff, E. & Jefferson, G. A simplest systematics for the organization of turn taking in conversation. *Language*, 50, 676-735, 1974.

[17] Savidis, A., Stephanidis, C., Korte, A., Crispien, K. & Fellbaum, K. A generic direct-manipulation 3D-auditory environment for hierarchical navigation in non-visual interaction. *Proceedings of Assets '96*. ACM Press, New York, NY, 1996.

[18] Shneiderman, B. Direct manipulation. *IEEE Computer, 16*, 8, 57-69, 1983.

[19] Shneiderman, B. Direct manipulation for comprehensible, predictable and controllable user interfaces. *Proc. of the 2nd international conference on intelligent user interfaces*. ACM Press, NY, 1997.

[20] Suchman, L. *Plans and situated actions*. CUP, Cambridge, 1987.

[21] Vanderheiden, G. Fundamental principles and priority setting for universal usability. *Proceedings of CUU '00*. ACM Press, NY, 2000.

[22] Winberg F., and Hellström, S.-O. Qualitative aspects of auditory direct manipulation. *Proc. of the 7th International Conference on Auditory Display*. Helsinki University of Technology, Finland, 2001.

[23] Yu, W., and Brewster, S. Multimodal virtual reality versus printed medium in visualization for blind people. *Proceedings of Assets '02*. ACM Press, NY.