KUNGL
TEKNISKA
HÖGSKOLAN

CID-132 • ISSN 1403-0721 • Department of Numerical Analysis and Computer Science • KTH

Qualitative Aspects of Auditory Direct Manipulation: A Case Study of the Towers of Hanoi

Fredrik Winberg & Sten-Olof Hellström

CID, CENTRE FOR USER ORIENTED IT DESIGN

Fredrik Winberg & Sten-Olof Hellström

Qualitative Aspects of Auditory Direct Manipulation: A Case Study of the Towers of Hanoi
Report number:CID-132
ISSN number: ISSN 1403-0721 (print) 1403-073X (Web/PDF)
Publication date: July-August 2001
E-mail of author: fredrikw@nada.kth.se, soh@nada.kth.se

Reports can be ordered from:

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE-100 44 Stockhom, Sweden
Telephone: + 46 (0) 8 790 91 00
Fax: + 46 (0) 8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

QUALITATIVE ASPECTS OF AUDITORY DIRECT MANIPULATION A CASE STUDY OF THE TOWERS OF HANOI

Fredrik Winberg

Royal Institute of Technology
Centre for User Oriented IT-Design
Lindstedtsvägen 5, SE-100 44 Stockholm, Sweden
fredrikw@nada.kth.se

Sten Olof Hellström

City University
Department of Music
London EC1V 0HB, UK
soh@nada.kth.se

ABSTRACT

This paper presents the results from a qualitative case study of an auditory version of the game Towers of Hanoi. The goal of this study was to explore qualitative aspects of auditory direct manipulation and the subjective experience from playing the game. The results show that it is important to provide a way of focusing in the auditory space. Articulatory directness was also an important issue and feedback should support the movement of the objects in the auditory space.

1. INTRODUCTION

This work is concerned with auditory direct manipulation and how to make graphical user interfaces with direct manipulation accessible for blind computer users.

Direct manipulation is a fundamental concept within human-computer interaction (HCI) and most graphical user interfaces rely on this concept. Some claims of direct manipulation are that it is easier to learn basic and new advanced features, it improves the user's confidence in using the system and it encourages usage by being more enjoying to use [1]. However, the notion of direct manipulation is not present in modern screen readers that many blind computer users use. The information in these applications is presented in a linear way using either speech synthesis or a Braille display, and does not allow the concurrent and spatial nature of the information to be displayed.

Direct manipulation is based on the following properties: [1, 2]

- *Continuous representation of the object of interest.*
- *Physical actions or labeled button presses instead of complex syntax.*
- *Rapid incremental reversible operations whose impact on the object of interest is immediately visible.*

This means, for example, that when moving a file instead of typing the command on your keyboard or choosing from a list of actions, you simply point your mouse at the file you want to move, grab it by pressing down the mouse button, drag it to the place you want it to be and drop it by releasing the button.

One important aspect of direct manipulation is continuous presentation, which means that the objects presents themselves in a continuous manner, the user does not need to browse and look for the objects. A crucial result of a direct manipulation interface is that it supports the recognition of objects, for example the use of menus or buttons helps the user to recognize the name instead of forcing the user to memorize the exact name and the exact syntax of the command of interest.

1.1. Previous research

Previous research in direct manipulation interfaces for blind computer users has included both using tactile devices together with audio [3, 4] and the use of 3d audio and a data glove [5]. But these have not addressed the importance of continuous presentation. In these examples only a subset of the interface objects are presented and the user has to browse the auditory space in order to get an overview. This is rather interacting directly with interface objects than direct manipulation as defined and used within HCI.

Interaction with complex auditory spaces has been reported by for example Saue [6], where a framework for interaction with spatial data is presented. The user walks around the data using the mouse and both local and global information is used to present the data and guide the user.

Another example of interaction with auditory objects is the work by Pitt and Edwards [7], where the cursor is used as a virtual microphone when exploring and choosing items from a menu.

1.2. Research questions

The two main research questions we want to address are

- Is auditory direct manipulation possible?
- Is auditory direct manipulation interesting or do we have to seek other paradigms for interaction with an auditory interface?

In order to answer the above questions we have implemented an auditory version of the game "Towers of Hanoi", and performed two user studies on this game.

2. AUDITORY TOWERS OF HANOI

In the game "Towers of Hanoi", you have three towers and three or more discs that are placed on the left tower. The object of the game is to move the discs from the left to the right tower. You can place a disc on any of the three towers, just as long as you do not move a larger disc on top of a smaller one.

2.1. Disc identity

The sonification model is exclusively based on the sounds of the discs. Every disc has a sound that differs mainly in pitch and in timbre. The larger the disc, the lower the pitch. The sounds are slightly mistuned with respect to each other to get a better separation (cf. [8, pp. 490-493]).

There is a large gap in pitch between every other disc, which groups the discs two by two. Within each group the main difference between the sounds is the timbre that originates in the use of either harmonic or inharmonic spectra.

2.2. Disc location

In order to distinguish which tower a disc is located on, both stereo panning and amplitude envelopes are used. The most obvious is to use stereo panning, left, middle and right. This is however not sufficient since it could be hard to hear the difference when using multiple complex sounds. Because of this, we have also used different amplitude envelopes. If a disc is placed on a tower to the left or to the right, the envelope has a very short attack and a long decay. If a disc is placed on the middle tower the attack and the decay have the same length.

A disc's vertical position is represented by the length of the sound, the higher the disc is placed the shorter the sound. The length of the sounds varies between 238-900 ms.

2.3. Mouse interaction

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. This is similar to the work presented in [7], where the mouse pointer is used as a microphone. In this implementation we use the volume difference only in discrete steps, either the cursor is on or off a tower, the distance is not mapped to the volume continuously. There are also short transition tones that tells the user when the cursor is moved from one tower to another.

The mouse could be seen as a virtual microphone, as in [7], that focuses on one tower at the time. By doing this, the user can direct the attention to a subset of the auditory space. You could also see this as a rather primitive way of auditory zooming [6, 9], using just an increase in volume instead of increasing the complexity of the display.

The sonification model is described more thoroughly in [10].

3. THE FIRST STUDY

The object of the first study was to investigate the nature of continuous presentation and what this could mean in an auditory interface. We also wanted to find out whether auditory direct manipulation was at all possible to achieve.

This experimental study compared three different presentation modes: parallel (all sounds repeating in parallel), serial (in sequence one sound at the time), and overlapping (in a rapid sequence with a slight overlap) when playing the game with three and four discs. The question was what was continuous enough and how the change in continuity would affect performance.

Given the limited complexity possible for all subjects to complete the game Towers of Hanoi (four discs in three different locations), the study showed that the presentation mode gave no significant differences in neither time to complete nor number of errors. When the subjects were asked which presentation mode they preferred and which they thought they performed best with, most subjects said the overlapping mode.

These results fail to support previous research that suggests that complex information should not be presented using continuous sounds [11]. Different explanations for these results include issues of complexity (to notice a difference between these presentation modes the complexity must be higher), navigation (the way the user browse and interacts with the display facilitates the use of continuous sounds), and sonification model (the sonification model is robust enough to being able to present these sounds in a continuous manner).

This experiment is described in detail in [10, 12].

4. THE SECOND STUDY

The goal of the second study, presented in this paper, was to explore qualitative aspects of auditory direct manipulation and the subjective experience from playing the game. The version used in this study was the overlapping version.

4.1. Procedure

Each session was divided into two parts. In the first part, the subjects learned about the game "Towers of Hanoi" and how to play it using the auditory version. The subjects were able to try three, four and five discs, but were not forced to try or complete all of these.

The second part was the actual interview, where the subject was asked to describe what it was like to play the game. The questions included issues like

- How would you describe the experience from playing this game?
- How would you describe the objects you are moving?
- Does it feel like you are moving actual objects or does it feel like something else?
- How easy is it to get an overview? Do you feel that the objects are there all the time or are they presented one by one?
- How would you make this game better?
- How could this way of interaction be used in your daily use of a computer?

4.2. Subjects

There were four subjects in this study and they were all blind. All the subjects had prior knowledge of the game "Towers of Hanoi".

- Subject 1 (male, 49) works as a computer expert at a company producing and selling assistive devices and rates his computer knowledge as advanced.
- Subject 2 (female, 39) works as a switchboard operator and rates her computer knowledge as average.
- Subject 3 (male, 54) works as an administrator at a handicap organization and rates his computer knowledge as advanced. He participated in the first study of the auditory Towers of Hanoi.
- Subject 4 (male, 30) works as an IT-technician at a handicap organization and rates his computer knowledge as advanced. He also participated in the first study.

5. RESULTS

One interesting observation during this study was that it yielded a lot of qualitative data while the subjects were playing the game. Without being asked to, the subjects described what they were doing, just like a thinking aloud-session. When asked about this, Subject 3 said that he had never thought about it before, but it could be that you as a blind computer user are used to describing what is happening on the screen. Very often the person standing next to you can not see what is presented on the Braille display or hear what is said by the speech synthesizer like when using a regular computer screen, so the only way to communicate what is happening is to describe it aloud.

5.1. Elements of direct manipulation

5.1.1. Continuous presentation

One essential aspect of direct manipulation is the notion of continuous presentation [2]. Continuous presentation means that all the interface objects should be present all the time, the objects should present themselves to the user rather than forcing the user to search for the objects. This requires that the interface gives the user an overview and constant presentation of the interface objects.

All subjects reported that they thought that the sounds were present all the time, even though they were played in sequence. Subject 4 described this as "they [the sounds of the discs] are played one by one, but you still think that you get like an overview, where they are".

Another aspect of this was described by Subject 1 as "… [the fact that] the sounds are there all the time is also dynamic … it goes on … you can't stop and everything becomes quiet and you can think for a while … it is affording, they are still there waiting for something more to happen". This captures the essence of continuous presentation, the objects should be available for interaction and present themselves without any need for the user to ask for it.

The focus feature of the mouse was used by the subjects in order to get a better understanding of where the discs were placed. Subject 1 described it as in order to understand the whole auditory space, he "walked around with the mouse from tower to tower" and listened to the different discs.

The focus feature was also used to get a better overview of the auditory space, it made it possible to concentrate on a specific tower without losing track of the other two towers. Subject 3 said that "even if it focuses rather heavily on sound, on the volume, you still have an overview of the three towers and that is good since it is very important … to have an overview all the time".

For subject 2, the focus feature seemed to be the one thing that made playing the game at all possible: "it is a little bit hard in the beginning, it is very confusing, you hear everything at the same time … until you realize that you can move and listen to one tower at the time. … if you listen to them at the same time you go nuts, you have to go to the towers and listen".

This suggests that in order to have continuous presentation, there should be support for discriminating the different sounds and some sort of focus or zooming on specific parts without losing track of the other parts.

5.1.2. Physical actions

All subjects thought that the way they moved the discs was intuitive and very direct. Subject 1 said that "as soon as I decide to move a disc and press the left mouse button I have it … whatever I do, something happens immediately following that".

One important aspect of the property of physical actions in direct manipulation is articulatory or mimetic directness. Articulatory directness could be described as the relationship between how you actually physically perform an action and what it means to the system, for example that any action with the input device should mimic the desired change on the interface object [2].

Since the movement with the mouse mimics the real movement of the discs, a movement from left to right with the mouse corresponds to a movement in the interface from left to right, none of the subjects had any problems at all understanding how to move the discs and what happened if you were to move the mouse from left to right.

Subject 3 said "it feels like moving around physical objects since you have a physical orientation in the room … I do have the feeling of moving some kind of real objects rather than just abstract objects".

This spatial mapping was something that was really appreciated by all subjects. Subject 2 explicitly commented on how she really liked the feeling of actually grabbing the virtual objects and moving them physically to the desired location, something that she knows is possible in graphical user interfaces and she had really missed doing herself.

It seems like the articulatory directness, the spatial mapping of the mouse movement, was a key to making the task more physical and direct rather than abstract and complex.

A feature that caused some initial confusion for Subject 4 was the fact that the vertical movement of the mouse makes no difference, you always pick up the disc on top. He said: "I wondered which [disc] I will move … but then it is obvious that you move the one on top". Even if this violates the physical metaphor, only one subject commented on this, and when

realizing that the vertical movement had no significance the subject had no problems at all adjusting to this.

5.1.3. Reversible actions with immediate feedback

The reversibility proved to be self-evident since all subjects on several occasions moved a disc to a location and then changed their mind and moved it back. Subject 1 showed this when moving by talking to the discs: "no, we can't move this anywhere … you have to go back".

When moving a disc the feedback is immediate and none of the subjects felt that they had to wait when they moved and dropped a disc on a tower. However, one drawback was that there was no continuous feedback when moving a disc, only discrete steps when moving from one tower to another. This made it hard for two of the subjects to know whether they were actually moving a disc or not.

5.2. Subjective experience

All subjects thought that this game was fun and stimulating to play even though they all found the game was quite complex. Subject 1 said that "the game itself is hard, to know how to move them [the discs] … you have to be systematic, it's a tricky game. … but it's fun!".

Despite that the game was hard, none of the subjects had any problems understanding the sonification model or playing the auditory game. All of the subjects could solve the game using three discs, three could solve it when using four discs and one was able to solve it when using five.

Subject 2 did have problems in the beginning and kept saying things like "these sounds are ugly", "very annoying sounds", and "what a jumble of sounds", but after a short while these comments stopped and the subject played the game silently. In the following interview the subject agreed that "it was hard in the beginning … it was a real mess and I thought that this will never work … but it was fun after a while". As described above, this change in attitude was explained by the subject as being related to the focus feature of the mouse, which made the "jumble of sounds" into something intelligible. After these initial problems, this subject was the only one who solved the game using five discs.

One thing that all subjects liked and commented specifically on was the use of stereo to spatialize the information. The use of stereo in screen readers is very limited and just one subject had actually tried a screen reader that used stereo.

5.3. Suggested changes

As mentioned before, specific feedback when moving a disc was a feature that some of the subjects thought was missing. One way of doing this could be to add some sort of sound effects when picking up, dragging and dropping an object. Subject 1 described this in a very animated way: "one sound effect you could add to this is when you drop a disc it could say 'ker-boff!' … then you really have dropped it … as well as when you press and lift [it could say] 'heave!' … [this would be] more fun and direct". All subjects thought that adding feedback on movement would make the game both more fun to play and easier to use.

Other comments touched upon things that would make the game perceptually more realistic, for example Subject 4 suggested that it would be more realistic if you could accidentally drop a disc between two towers. However, this was nothing that he believed would make it easier or better, just more complex and confusing.

5.4. Future applications

During the second part of the interview, the subjects were asked to give suggestions on what this kind of auditory interfaces could be used for. This was done both to get feedback on how useful the subjects thought that auditory direct manipulation was and to get ideas on future applications.

In general, the subjects liked this way of interacting with the computer and their comments were more about how much they liked it rather than any specific applications. Among the suggestions were moving text around in a word processor, navigation on a desktop and dragging files between folders, and getting a general overview of a large set of objects.

6. DISCUSSION

There are many questions remaining to be answered, for example if these principles are interesting when interacting with an auditory interface. These studies shows that it could be a good way, but what happens when the complexity increases? In the auditory Towers of Hanoi we have only used 5 discs due to the nature and the complexity of the underlying game (most subjects, both blind and sighted, simply could not play the original game using more than 4 or 5 discs). When implementing a desktop interface with files and folders the number of objects most certainly will be larger than this, how scalable is this approach in that case? Does it offer the blind computer any new benefits at all?

As mentioned above, the subjects came up with a number of possible situations where auditory direct manipulation could be used, and this will guide the next step in this research, applying auditory direct manipulation in a real context, not just in a game. The next step in this work is to implement drag & drop on a desktop with files and folders.

This study points at two important qualitative aspects of auditory direct manipulation: articulatory directness and focus. The use of a physical metaphor, in this case simply the use of stereo to spatialize the discs, increased the articulatory directness. This was very much appreciated by the subjects and made the task of playing the game more enjoyable and easier to understand. The subjects even thought it felt like they were moving actual, physical, objects. The fact that the sounds themselves were not metaphorical or based on real world sounds, as in auditory icons [13], did not seem to be of importance for the subjects.

The focus feature of the mouse proved to be essential in order to get an overview of the complex auditory space. The subjects could concentrate on a subset of the interface objects without losing track of the other objects. This could explain

the results in the first study where the differences in mode of continuous presentation proved not to be significant [10]. The focus also supports the articulatory directness, since there is a direct one-to-one mapping between the physical movement and the audio feedback. As stated above, more explicit feedback on movement is needed to make this mapping complete.

As a concluding remark we claim that auditory direct manipulation is both possible and has a potential of being an important improvement for blind computer users in their use of computers. Subject 2 gave a good example of this: "when there is a new application at my workplace, we get instructions like 'then you click here and you drag this one here', and I can't do this, I always need to have special commands … and no one knows these commands and can show me, unless I do it myself… perhaps this could help me". In this particular case, the subject does not believe that auditory direct manipulation will replace the use of keyboard shortcuts, or even be better, but she thinks that in some cases, like the one just described, it could make life a little bit easier.

7. REFERENCES

[1] B. Shneiderman, "Direct manipulation: a step beyond programming languages," *IEEE Computer*, vol 16, pp. 57-69, 1983.

[2] E.L. Hutchins, J.D. Hollan, and D.A. Norman, "Direct manipulation Interfaces," in *User centered system design*, D. A. Norman and S. W. Draper, Eds. Hillsdale, NJ, Lawrence Erlbaum Associates Inc, pp. 87-124, 1985.

[3] GUIB Consortium, *Final Report of the GUIB Project: Textual and Graphical Interfaces for Blind People*, London: Royal National Institute for the Blind, 1995.

[4] H. Petrie, S. Morley, and G. Weber, "Tactile-based direct manipulation in GUIs for blind users", in *Conference companion on Human factors in computing systems, CHI'95*, New York: ACM, pp. 428-429.

[5] A. Savidis, C. Stephanidis, A. Korte, K. Crispien, and K. Fellbaum, "A Generic Direct-Manipulation 3D-Auditory Environment for Hierarchical Navigation in Non-visual Interaction", in *Proceedings of Assets'96*, New York: ACM, pp. 117-123, 1996.

[6] S. Saue, "A model for interaction in exploratory sonification displays," in *Proceedings of ICAD 2000*, available at: http://www.icad.org/websiteV2.0/Conferences/ICAD2000/ICAD2000.html

[7] I.J. Pitt, and A.D.N. Edwards, "Navigating the Interface by Sound for Blind Users", in *People and Computers VI, Proceedings of the HCI'91 Conference*, D. Diaper, and N. Hammond, Eds. Cambridge University Press, pp. 373-383, 1991.

[8] A.S. Bregman, *Auditory scene analysis: The Perceptual Organization of Sound*, Cambridge, MA: MIT Press, 1990.

[9] U. Axen, and I. Choi, "Investigating geometric data with sound," in *Proceedings of ICAD'96*, available at: http://www.icad.org/ICAD96/proc96/axen.htm

[10] F. Winberg, and S.O. Hellström, "The quest for auditory direct manipulation: the sonified Towers of Hanoi," in *Proceedings of the 3rd International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT 2000)*, P. Sharkey, A. Cesarani, L. Pugnetti, and A. Rizzo, Eds. Reading, UK, pp. 75-81, 2000.

[11] W.W. Gaver, R.B. Smith, and T. O'Shea, "Effective sounds in complex systems: The ARKola simulation," in *Proceedings of CHI'91*, New York: ACM, pp. 85-90, 1991.

[12] F. Winberg, and S.O. Hellström, "Investigating Auditory Direct Manipulation: Sonifying the Towers of Hanoi," in *CHI 2000 Extended Abstracts*, New York: ACM, pp. 281-282, 2000.

[13] W.W. Gaver, "Using and Creating Auditory Icons," in *Auditory display: Sonification, audification, and auditory interfaces*, G. Kramer, Ed. Reading, USA: Addison-Wesley, pp. 417-446, 1994.