KUNGL
TEKNISKA
HÖGSKOLAN

CID-136 • ISSN 1403-0721 • Department of Numerical Analysis and Computer Science • KTH

Auditory Direct Manipulation for Blind Computer Users: Report on Work in Progress

Fredrik Winberg

CID, CENTRE FOR USER ORIENTED IT DESIGN

Fredrik Winberg

Auditory Direct Manipulation for Blind Computer Users: Report on Work in Progress

Report number:CID-136
ISSN number: ISSN 1403-0721 (print) 1403-073X (Web/PDF)
Publication date: August 2001
E-mail of author: fredrikw@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

International Symposium on Signal Processing and its Applications (ISSPA), Kuala Lumpur, Malaysia, 13 – 16 August, 2001. Organized by the Dept. of Microelectronics and Computer Engineering, UTM, Malaysia and Signal Processing Research Centre, QUT, Australia

AUDITORY DIRECT MANIPULATION FOR BLIND COMPUTER USERS REPORT ON WORK IN PROGRESS

Fredrik Winberg

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

ABSTRACT

This paper describes ongoing work with auditory direct manipulation for blind computer users. Two studies have been performed on an auditory version of the game Towers of Hanoi, and the results have pointed out important aspects of auditory direct manipulation such as focus in the auditory space and articulatory directness. Future research directions include studies on these aspects, collaboration between sighted and blind subjects and an implementation of auditory drag&drop in a file system.

1. INTRODUCTION

Direct manipulation is a fundamental concept within HCI (human-computer interaction) and is based on that the interface has the following properties: [3, 9]

- *Continuous presentation of all interface objects*
- *Physical actions*
- *Rapid incremental reversible operations with immediate feedback*

This means that you for example move a file by simply pointing the 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. If you are not satisfied with the action, you could very easily move the file back to its original place. During this whole process you get feedback of the file moving on the screen when you move the mouse. All objects that are available to you in any given moment are also displayed on the screen.

This direct and in many ways intuitive way of interaction has been very influential in today's graphical user interfaces and will influence the way we interact with computers for a long time. One serious drawback though is that this kind of interaction is not accessible at all for blind computer users, screen readers does not support direct manipulation. This means that, for example, even though blind people are using the same software as their sighted co-workers, they are practically using a different software with a completely different look and feel. Another problematic area is instructions; blind people could be left out when new software is introduced at the workplace when they do not understand how to use the new software since the instructions is based on direct manipulation interaction techniques. [11]

The three general research questions that we want to address in this project are:

- Is auditory direct manipulation possible?
- Could auditory direct manipulation support collaboration between sighted and blind users?
- Is auditory direct manipulation interesting or do we have to seek other paradigms for interaction with an auditory interface?

2. AUDITORY DIRECT MANIPULATION

Auditory direct manipulation is a rather uncharted territory both in research and development, given that we talk about direct manipulation as defined above and how it is used in modern graphical user interfaces, and not just interacting directly or almost directly with interface objects. Previous research has been concerned with giving blind computer users a more direct way of interacting with interface objects [2, 8], but has not dealt with direct manipulation itself. The main aspect that is missing in these cases is continuous presentation, only a subset of all the objects are presented at the time and the only way to get an overview of all objects is to manually browse the environment. This clearly does not follow the principle of continuous presentation.

Other work has been done on complex auditory interfaces (see for example [1]), but many of these has been monitoring tasks were the focus has been on the display of information rather than the interaction with auditory objects [7].

The key issues of our research on auditory direct manipulation are:

- It should support the principles of direct manipulation, especially continuous presentation of all interface objects
- It should support collaboration with users of the graphical counterpart, therefore it should support the same actions, the same concepts, and the same spatial layout

2.1 Continuous presentation

One essential aspect of direct manipulation is the notion of continuous presentation [3]. This means that all objects should be presented in a continuous manner. The user should

0-7803-6703-0/01/$10.00©2001 IEEE

not have to look for the interface objects, they should present themselves by being there all the time. This is probably the most difficult one to solve when using audio (see for example [8])

Since audio has a strong ambient quality and is not bound by line of sight, you cannot "listen away" or focus in an effective way like you can when looking [4]. When the auditory space gets more complex (i.e. when the number of objects increase), this gets harder. Still, it is one of the key principles and an essential part when implementing direct manipulation. Since presenting the sounds in a continuous manner is hard, the question is rather what is continuous enough. How do we deal with the ambient quality of sound without losing the overview and the feeling of the objects being there all the time, and how do we avoid overloading the auditory space without hiding some of the objects?

2.2 Physical actions

The second property, physical actions, could be seen as articulatory or mimetic directness [3]. This means that the action should mimic the desired change on the interface object, for example when moving a file from left to right on the desktop, the movement of the pointing device should also go from left to right. Interesting issues includes how to support spatial movement with the pointing device and placement of the interface objects, how to support precision positioning of the cursor [6], and the kind of pointing device that are used.

2.3 Rapid incremental reversible operations with immediate feedback

The final property, rapid incremental reversible operations with immediate feedback, requires that all operations could be reversed, for example moving an object back to it's original position after moving it to an undesired location, and that the new location and the new status of the auditory space is immediately audible for the user. The interface should also give immediate feedback on the movement itself, not just the beginning and the end [10, 11]. The main issue is how to give feedback on a spatial movement or action.

3. AUDITORY TOWERS OF HANOI

In order to study auditory direct manipulation, we have designed and implemented an auditory version of the game "Towers of Hanoi". In this game, the user interacts using a regular mouse and a pair of headphones, with no screen at all.

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. 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.

Figure 1 shows sonograms of four discs. Please note the large difference in frequency between the second and the third disc, as well as the differences in the disc's spectra.

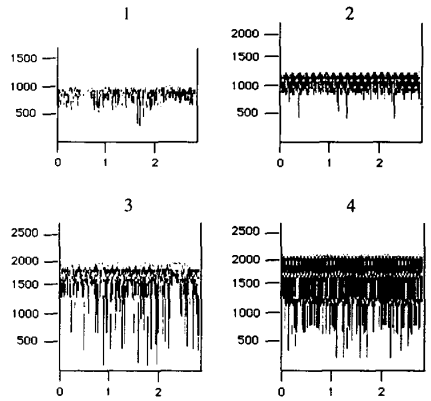

Figure 1. Sonograms of four discs from largest to smallest (time on the horizontal axis and frequency on the vertical)

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 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 [5], where the mouse pointer is used as a virtual microphone. In this implementation though, we use only discrete steps, the cursor is either on or off a tower, the distance between a tower and the cursor is not mapped. By doing this, the user can direct the attention to a subset of the auditory space. There are also short transition tones that tells the user when the cursor is moved from one tower to another.

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

3.1 The first study

The first study we did on this game was an experimental study where we investigated continuous presentation and what this could mean in an auditory interface.[10] We compared three different levels of continuous presentation where the main difference was the length and type of the intervals when repeating the sounds.

In this study, we tried to answer the question what was continuous enough. Is it sufficient to play the sounds in sequence with no overlap at all (the serial condition), or do we have to fully implement continuous presentation where all sounds keep repeating continuously and simultaneously (the parallel condition)? We also implemented a mixture of these first two where the sounds were played in sequence with a slight overlap (the overlapping condition).

The results from this study showed that with a limited set of auditory objects (4 discs in 3 different locations) the level of continuous presentation gave no significant difference. However, most subjects preferred the overlapping condition.

3.2 The second study

The second study was a qualitative case study where we wanted to explore the qualitative aspects of auditory direct manipulation and the subjective experience from playing the game.[11] The four subjects were all blind and had different computer experience, ranging from novice to expert user. The subjects first played the game and then responded to a semi-structured interview.

All subjects reported that they felt that the sounds were present all the time, even though they were played in sequence. The way that the subjects used the focus feature of the mouse seemed to support this feeling of continuous presentation. This could be a very important finding on how to deal with continuous presentation in an auditory interface, but requires more experiments.

The focus feature of the mouse was used by the subjects in order to get a better understanding of where the discs were placed and to get a better overview of the auditory space. This feature made it possible to concentrate on a specific tower without losing track of the other two towers. It was thought to be very important for the subjects and was the most appreciated part of the auditory interface.

All subjects thought that the way they moved the discs was intuitive and very direct. Since the movement with the mouse mimics the real movement of the discs, none of the subjects had any problems at all understanding how to move the discs. This supported the articulatory directness of the interface and was greatly appreciated by all subjects.

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.

The subjects thought that sufficient feedback on grabbing, dragging and dropping was missing. This would enhance the interaction and make it feel more direct and engaging.

4. DISCUSSION

The work so far has showed that auditory direct manipulation is indeed possible. The studies on the auditory version of Towers of Hanoi has given valuable pointers on how to implement certain features of auditory direct manipulation, especially on focus in the auditory space and articulatory directness.

There are many questions remaining to be answered, for example if these principles really are the best way for interaction with an auditory interface. These studies shows that it could be a good way, but what happens when the complexity increases? In the Towers of Hanoi we have only used 5 discs due to the nature and the complexity of the underlying model, most subjects simply could not play the game using more than 4 or 5 discs, neither audio-only nor visual/haptic.

Additionally, the Towers of Hanoi is just a game with little or no relation to auditory direct manipulation in the end user's daily use of computers. The use of only three discrete locations is hard to maintain when implementing for example a desktop system, since there are many more locations that needs to be represented.

Future research will involve the design, implementation and evaluation of an application of auditory direct manipulation. This will be an implementation of drag&drop in a file hierarchy, a task that is more common to a real situation for a blind user than perhaps playing a game is.

In order to answer the question whether auditory direct manipulation could support collaboration between sighted and blind users, an experiment is being planned where one blind and one sighted subject are supposed to solve the Towers of Hanoi collaboratively taking turns moving the discs. Interesting questions in this study will be how the subjects talk about what they are doing, if they are using the same metaphors and concepts, and how the strategy is agreed upon between the subjects.

We believe that auditory direct manipulation is a promising way to give blind users better access to modern graphical user interfaces. The benefits of this are primarily that the blind computer user gets a more direct and more intuitive way of interacting with a complex auditory space, and that collaboration between sighted and blind computer users facilitates.

5. REFERENCES

[1] W.W. Gaver, R.B. Smith, and T. O'Shea, "Effective sounds in complex systems: The ARKola simulation", in *Conference proceedings on Human factors in computing systems, CHI'91*, pp. 85-90, 1991.

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

[3] 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., Lawrence Erlbaum Associates Inc, Hillsdale, NJ, pp. 87-124, 1985.

[4] G. Kramer, "An introduction to auditory display", in *Auditory display: Sonification, audification, and auditory interfaces*, G. Kramer, Ed., Addison-Wesley, Reading, USA, pp. 1-77, 1994.

[5] 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.

[6] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa, "A framework and testbed for studying manipulation techniques for immersive VR", in *Proceedings of the ACM symposium on Virtual reality software and technology, VRST'97*, pp. 21-28, 1997.

[7] S. Saue, "A model for interaction in exploratory sonification displays", in *Proceedings of the International Conference on Auditory Display, ICAD 2000* [online proceedings], URL http://www.icad.org/websiteV2.0/Conferences/ICAD2000/ICAD2000.html , 2000.

[8] 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 the second annual ACM conference on Assistive technologies, Assets '96*, pp. 117-123, 1996.

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

[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] F. Winberg, and S.O. Hellström, "Qualitative Aspects of Auditory Direct Manipulation: a Case Study of the Towers of Hanoi", accepted for publication at *The Seventh International Conference on Auditory Display ICAD 2001*.