KUNGL
TEKNISKA
HÖGSKOLAN

CID-74 ● ISSN 1403-0721 ● Department of Numerical Analysis and Computer Sience ● KTH

Investigating Auditory Direct Manipulation: Sonifying the Towers of Hanoi

Fredrik Winberg & Sten-Olof Hellström

CID, CENTRE FOR USER ORIENTED IT DESIGN

Fredrik Winberg & Sten-Olof Hellström

Investigating Auditory Direct Manipulation: Sonifying the Towers of Hanoi

Report number:CID-74
ISSN number: ISSN 1403-0721 (print) 1403-073X (Web/PDF)
Publication date: April 2000
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

Investigating Auditory Direct Manipulation: Sonifying the Towers of Hanoi

Fredrik Winberg
Centre for User Oriented IT-Design
Royal Institute of Technology
SE-100 44 Stockholm, Sweden
+46-8-790 9273
fredrikw@nada.kth.se

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

ABSTRACT

This paper presents a study of an auditory version of the game "Towers of Hanoi". In this study, we have compared three different strategies for continuous presentation of the objects. The focus for this study is to investigate the nature of auditory direct manipulation, where continuous presentation is one of the key aspects. The results show that the differences between the strategies need to be explored further by experimentation. Additionally, much effort has to be put on the learning phase of the auditory interface and the mouse interaction has to be investigated further.

Keywords

Auditory interface, direct manipulation, sonification model, blind users

INTRODUCTION

The use of sound in human-computer interfaces — especially when using sound not only as auditory cues in a visual interface, but as the only means of output — is a rather uncharted territory within the CHI community (see for example the proceedings of CHI'99).

One major difference between auditory screen reading software for blind computer users and ordinary graphical user interfaces is the notion of direct manipulation (see for example Hutchins, Hollan & Norman 1985). Direct manipulation is based on three important general properties: *continuous representation* of interface objects, *physical actions* instead of complex syntax, and *rapid incremental reversible operations* with *immediate feedback*. This is a paradigm that almost all graphical user interfaces today are based on. When trying to translate these interfaces to auditory interfaces that are accessible by blind users, the above properties makes this hard.

The general question we want to address is what direct manipulation means in an auditory interface. Is it at all possible?

Short paper presented at CHI 2000 Conference on Human Factors in Computing Systems, The Hague, The Netherlands, April 1-6, 2000

The particular focus of this study is on continuous presentation and how, if possible at all, this could be solved in an auditory interface.

TOWERS OF HANOI

The object of the game "Towers of Hanoi" is to move a number of discs from the leftmost to the rightmost tower using just three towers. You can place a disc at any of the three towers, just as long as you don't move a larger disc on top of a smaller one. The minimal number of moves to complete an *n*-disc Towers of Hanoi is 2^n-1, which of course limits the number of discs used.

We chose the game "Towers of Hanoi" for three reasons; (1) we wanted to have an actual game that could be seen as a challenge to solve, (2) the rules of the game are fairly easy to learn and the strategy is straightforward and doesn't change when increasing the number of discs, just the number of steps in the solution path, (3) it's easy to show the subjects a wooden model of the game in order for them to learn how to solve the game.

PROCEDURE

In this study we have used twelve subjects, of which one were blind. At the beginning of each session, the subject learns how to solve the game by using a wooden model of the game. This learning period isn't part of the study, but rather a way for the subject to get familiar with the game and learn how to solve it for three and four discs (a pilot study indicated that many subjects would not be able to complete the game with more than four discs in a reasonable amount of time). By doing this we are also trying to even out differences in prior knowledge of the game and get all subjects to have a useful model in mind when solving the auditory version of the game. Additionally, a wooden model is accessible by both blind and sighted subjects. The subject is then presented a randomized sequence of combinations of number of discs and presentation strategies (see section Sonification Model below) and is asked to solve them one at a time.

SONIFICATION MODEL

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 an even better separation (Bregman, 1990). There is a large gap in frequency between the second and the third disc, which groups disc one and two together and disc three and four together. Within each group the main difference between the sounds is the timbre that originates in the use of either harmonic or inharmonic spectra.

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 release. If a disc is placed on the middle tower the attack and the release 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.

We have used the mouse as input device. The volume of the discs on the tower that the mouse cursor is located on is increased. If the subject moves the mouse to another tower, the volume of the discs on this tower will increase and the volume of the discs on the previous tower will decrease.

The major goal of this sonification model is to explore the nature of continuous presentation and what this could mean in an auditory interface. We have tried the two extreme cases for continuous presentation. The first is where all the sounds keep repeating simultaneously. The second is where all sounds are played in sequence with no overlapping at all. We have also tried a mix between these two where the sounds overlap slightly. The order of the sounds in the last two cases is from the largest to the smallest disc.

RESULTS

Most subjects preferred the overlapping version, either from personal preferences or from reasoning (two subjects actually thought that the different strategies were all the same but they judged the three strategies based on what they thought others would think). Many subjects also thought they performed better with the overlapping version. This was true in the beginning of the session, but after completing more conditions the difference in performance levelled out.

Even though nine of the subjects did understand and achieved well in the experiment, three didn't understand at all and consequently could not complete all the conditions. One explanation that was posed by one of these subjects was that the instructions were insufficient and that the subject didn't really understand the sonification model completely. This was further emphasized by other subjects that did complete the game. This is not so surprising since most people don't have any prior knowledge or experience of auditory interfaces.

The mouse interaction posed more problems than we expected. The major difficulty that the subjects had during the whole study was to hear where the mouse cursor was. The middle tower was often very hard to find and a very common scenario was the subject flipping from left to right without ever finding the middle. Just increasing the volume to show the location of the mouse was not sufficient. A solution to this problem could be a short tone that tells the user when the mouse moves from one tower to another.

DISCUSSION

The results of this study will be used as valuable inputs to a future experiment where we will examine the features of auditory direct manipulation in depth.

This study also indicates that auditory direct manipulation is not only possible but it is also an effective way of interacting with a complex auditory space. This also gives blind computer users a completely new way of interacting with a computer, a way that so far only has been possible when using a graphical user interface.

REFERENCES

Bregman, A. S. (1990). *Auditory scene analysis: The Perceptual Organization of Sound* (pp. 490-493). Cambridge, MA: MIT Press.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation Interfaces. In D. A. Norman & S. W. Draper (Eds.), *User centered system design* (pp. 87-124). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.