INTERACTING WITH SONIFICATIONS: AN EVALUATION

Sandra Pauletto[1,2] and Andy Hunt[2]

[1]School of Computing and Engineering,
The University of Huddersfield, Queensgate,
Huddersfield, HD1 3DH, UK
`s.pauletto@hud.ac.uk`

[2]Department of Electronics, The University of
York, Heslington, York, YO10 5DD, UK
`adh@ohm.york.ac.uk`

ABSTRACT

The aim of the experiment described in this paper is to evaluate and compare three different methods for interacting with an algorithm for the sonification of data streams. The experiment was carried out using an existing Interactive Sonification Toolkit as a high fidelity prototype. The experiment focused on measuring and comparing the efficiency and effectiveness of three interaction methods which differ in the degree of real-time control allowed to the user. Subjects were also asked to answer a questionnaire which gathered information about their perception of using the different interaction methods.

The experiment shows that the method providing the lowest degree of real-time control to the user is the least efficient. This method is also perceived to be the least pleasant, fast, clear and intuitive. There are no significant differences in terms of effectiveness and efficiency for the remaining two methods both in terms of objective measures and user perception. Finally the method allowing a medium degree of control to the user is judged to be significantly more pleasant than the others.
[Keywords: interaction, sonification, evaluation]

1. INTRODUCTION

Interest in interaction - moving beyond the simple triggering of sonified sounds - has been shown by the sonification community since the early 1990s. The mouse has been the first choice of interface for many researchers working in interactive sonification as it is the most common computer interface in use to this day. Winberg and Hellström [2] used the mouse as a virtual microphone and in Brazil et al [3] multiple musical tunes are navigated using the mouse. Hermann used it to interact with data spaces [4] in his early examples of model-based sonifications. In recent years, with the higher processing power of modern computers, more research on new interfaces can be found. Haptic interfaces were explored by DiFilippo in [5] and Beamish et al [6]. Hermann and colleagues have explored the use of many novel interfaces such as the 'gesture desk', the 'audio-haptic ball', etc. Recent studies have also evaluated the use of some interfaces such as the mouse [7], the keyboard [8] and the tablet [9] for the navigation of two dimensional data sets. Finally, an excellent collection of papers on interactive sonification can be found in the special issue of IEEE Multimedia [10] which was dedicated, in 2005, to interactive sonification. A signal of the growing the interest on interactive sonification is the instigation of the Interactive Sonification workshop which first took place in Bielefeld in 2004 and then in York in 2007 [11]. The work presented here aims to shed some light on how data analysis can be improved by allowing interaction with a sonification algorithm.

2. EVALUATING THE INTERACTION

The goal of a software tool which uses sonification to display complex data sets is to support data analysis and exploration efficiently and effectively. A fundamental task in data analysis is to be able to identify particular structures present in the data under examination.   Three separate factors can affect the efficiency and effectiveness of data analysis via sonification:
  1) the specific data set used;
  2) the sonification algorithm;
  3) the interaction method.

This means that once a sonification method has proven to be a good display for a particular data set, and once the user knows what kind of data structures can be found in the data set, the efficiency and the effectiveness (with which a user analyses a data set and recognises the data structures it contains) depend only on how the user is allowed to *navigate* the auditory display. In this experiment the independent variable is the interaction method which is evaluated under three different conditions. The dependent variables are the time spent to complete a task (measure of efficiency: the higher the time, the lower the efficiency) and the number of incorrect identifications of data structures made during the execution of the task (measure of effectiveness: the higher the number of incorrect answers, the lower the effectiveness).

2.1 Experiment description

In this experiment, the subjects were asked to navigate and listen to a sonification using three different interaction methods. Their task was to recognise which data structures were present in the data set and in which order. Before starting the test, the subjects were trained to listen to examples of the sounds made this sonification method and to the particular kind of data sets provided in the test so that they gained experience in how to analyse and recognise the structures present in the data sets. This training was given so that the experimenter did not need to consider the sonification method and the type of data sets as variables in this experiment.
In the test, the subjects used the Interactive Sonification Toolkit developed by the authors and described in [1]. In this toolkit, the user is presented with a red rectangular area (or screen area, see

Figure 1) on the top half of the screen. This screen area represents the sonification from beginning to end (the beginning is mapped to the left corner and the end to the right corner of the area). The user can navigate the data by interacting with this screen area using two types of interfaces: the mouse and the shuttle interface [12] (see Figure 2).

2.2. The interaction methods

The three interaction methods studied in this experiment are:

1) Low interaction method

This method is included as it represents the most common way of interacting with data – using the mouse and menu options to set up user requirements, then playing the sound. In this method, the subject *selects* a section of the sonification by right-clicking the mouse somewhere within the screen area, dragging it towards the right, and then letting go of the right button of the mouse wherever the selection ends (Figure 1).
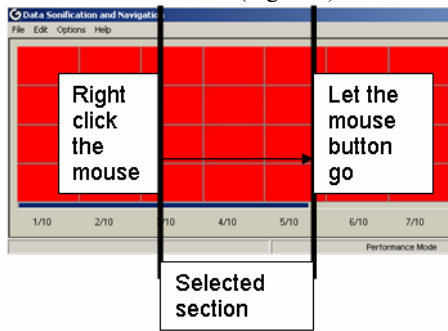

Figure 1: *Selection of section of sonification.*

After the selection has occurred, the subject needs to enter the amount of time in which he/she wants to hear the selected section of data. Finally, the subject clicks the button 'play' to hear the selected section. In this case the only possible manipulation is to choose the length of time in which to hear the selection (which could be considered a kind of temporal zoom). Sonifications are often presented and analysed using this method (i.e. by listening to the whole sonification at a set duration) and therefore this is the obvious method to which compare more interactive modes.

2) Medium interaction method

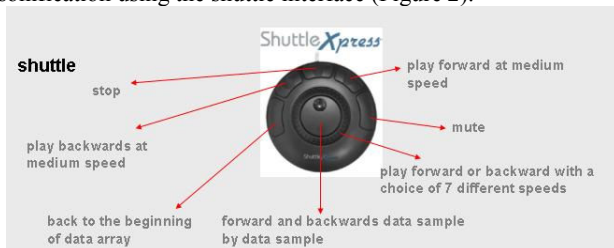In the second method, the subject is asked to navigate the sonification using the shuttle interface (Figure 2).

Figure 2: *The Shuttle XPress Interface*

The buttons and the wheels of the shuttle are mapped to defined presets as shown in Figure 2. With this interaction method, the subject can move backwards and forwards in the data at various constant speeds. He/she can jump to the beginning by pressing one button and can stop and change direction instantly.

3) High interaction

In the third method, as the subject left-clicks the mouse and drags it around the screen area, the sonification plays. By moving the mouse around in the screen area, the mouse is instantly mapped to the scaled values of the data set which are instantly fed into the sonification algorithm that produces the sound. The speed at which the data is played depends on the speed of the movement of the mouse on the screen. In this case the speed is rarely constant.

3. THE DATA AND ITS SONIFICATION

The structures present in the data need to be fixed and known by the experimenter so that, when the test subjects are asked to recognise the structures present, the correct answers can be counted.
It was a difficult task to decide what types of data structure to use in this experiment. Complex data sets come from both man-made systems and the natural world and they present us with an infinite variety of possible data structures. In this experiment only a few structures could be used and they needed to be controlled, i.e. completely known. There is no right way of deciding how to choose and create these structures. On the basis of the experimenter's experience in working with data sonification, five main data structures were considered to be very basic and common in data sets produced by any type of process (e.g. natural, mechanical, etc.):

1) a noisy structure;
2) a constant structure;
3) a linear structure (in particular an ascending linear ramp);
4) a discontinuous structure;
5) a periodic structure.

Each data set used in this experiment included all of these structures and each data set channel contained the same number of data samples (44100).
This number of data per channel was chosen to set the following timing reference: if the data was played back at audio rate (44.1kHz), then the sound would last one second.

3.1. The sonification algorithm

It was decided to map the data sets to the amplitude of a sine oscillator of fundamental frequency 261.6Hz (middle C).
This type of mapping is simple enough so that people can very quickly learn how the different data structures sound using this mapping.
•   The noisy structure sounds like noise.
•   The constant structure sounds like a sine wave of constant volume.

• The linear ramp sounds like a sine wave increasing in volume.

• The discontinuous structure sounds like a series of clicks and silences in between clicks.

• The periodic structure sonification results in a typical case of Amplitude Modulation. If the playing speed is low, it sounds like a kind of vibrato, while if the speed is high two sidebands typical of AM synthesis form and therefore we hear three pitches.

For this experiment it is important that people can easily recognise the data structures if they are presented with a simple data set. However, in order to be able to measure the effects of different interaction methods on the identification of the structures, it is important to create data sets that need repeated listening to be understood: if repeated listening was not needed, the action of navigating the data would not be needed and obviously would not be measurable.

The strategy used to make the datasets at the same time simple, but requiring repeated listening, was to construct them in such a way so that they would challenge the subject's hearing attention all the time. Experiments show [see 13; pp. 207-8] that if two different sequences of words are presented to subjects, one in the left ear and the other in the right ear, and after the subjects are asked to repeat one sequence, they usually cannot report the words heard in the non-attended ear. This led to the idea, in this experiment, of playing different streams in the two ears simultaneously. Usually subjects would need repeated listening to switch attention from the left to the right ear to recognise all the elements in the two streams. For this reason, it was decided that each data set should contain two channels of data, one panned to the left and one to the right, each containing two different sequences of the five structures mentioned above.

It was also decided that the different sections of structures should last different lengths of time (so that structures would not change simultaneously both in the left and in the right ear). Each channel had a sequence of 10 structures. Two sections of data containing the same structure could not be presented one after the other. Three different data sets were constructed. In each of them the order of the structures' sections and their length was different. Figure 3 shows a schematic representation of a typical data set used in this experiment.
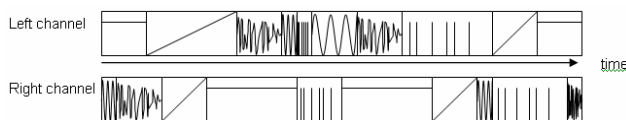

Figure 3: *Graphical representation of a data set used in the experiment*

4. THE EXPERIMENTAL PROCEDURE

4.1. Experimental design

This experiment has a within-subjects (or related) design. This means that the same group of subjects does the experiment under all the conditions (the three conditions correspond to the three different interaction methods).

Various aspects of the experiment are randomised so that effects due to order of presentation are eliminated:

• in order to eliminate errors due to the order of the conditions, each subject is presented with the conditions in a different order;

• every time the interaction method, i.e. the condition, changes, the data set also changes (otherwise the subject would already know the order of the data structures);

• finally, it is important not to always assign the same data set to one interaction method because this can cause errors: for example, one interaction method could be particularly good when used in conjunction with one particular data set.

The number of possible data set/condition combinations that follow the above rules can be calculated to tell us how many test subjects are necessary for the test. Let us call the three different interaction methods a, b, and c, and the three different data sets 1, 2 and 3. The total number of permutations without repetitions for the three interaction methods is 6 (abc, acb, bac, bca, cab, cba) and the total number of permutation without repetitions for the data sets is also 6. This means that in total we can have 6x6=36 combinations of data sets and conditions. However this can be halved if we eliminate some of the combinations such as a1b2c3, a1c3b2 which have the same data set/condition only in a different order. Finally 18 combinations were used and therefore 18 subjects were required. An a priori power analysis test confirmed that this number of subjects could detect a large effect size.

4.2. The test subjects

The average age of the subjects was 28. It was assumed that there would no be differences in judgment due to gender, and that, given a higher education level as background, differences in cultural background could also be disregarded. The subjects were fifteen male and three females. All the participants were British, apart from a Malaysian and a French person. The subjects were all researchers, students or lecturers of York University's Electronics Department. Sixteen subjects specialise in Music and Audio Technology. Sixteen subjects normally work with sound, while two do sporadically. The test was carried out in a silent room (in the recording studio performance area at the University of York, UK). Good quality headphones (DT 990 Beyerdynamic) were used with a wide frequency response ($5 - 35,000$Hz). This minimised the errors that could be due to external sounds and maximised the quality of sound reproduction.

4.3. Description of the experiment to the subjects

The task was explained to the subjects by the experimenter and the subjects had as much time as needed to familiarise themselves with the sonifications, the data structures and the interaction methods (typically they used about 5 minutes). The experimenter made sure that the subjects could easily recognise the data structures by giving a simple training that used a simple data set example. The test subjects were asked to navigate this simple sonification example using the three interaction methods and were asked to identify the structures present in the data set example. The data set consisted of one channel of data. This channel was sonified and sent to both left and right ears. The data set was divided into five sections, each one containing one data structure. The subjects started the experiment only when they felt

confident in recognising the structures and in using the interfaces.

4.4. The test

The experimenter uploaded and scaled appropriately one data set and panned the sound appropriately. The subjects were asked to write down the sequences of structures they heard in both ears just by writing the following letters to indicate the structures:

    N noise
    C constant
    L linear ramp
    D discontinuities
    S sinewave

The experimenter measured how long the subject took to recognise the sequences using a hand held chronometer. After having done the task using the three different interaction methods and the different data sets, each subject was asked to fill in a questionnaire. In the questionnaire the subject was asked to rate from 1 to 5 the pleasantness of each interaction method, the intuitiveness, the clarity and the quickness. The questionnaire was designed to gather subjective information about the perceived efficiency and effectiveness of the interaction methods which could be compared with the objective results produced by the test. In particular, the perception of quickness was considered to be linked with efficiency and the perception of clarity with effectiveness. Finally, intuitiveness was considered to be a measure of good affordance and close interaction feedback (qualities which should increase effectiveness) and pleasantness just a measure of how likeable an interaction method was. Then the subjects were asked to select their preferred interaction method and to comment on why they chose it.

5. EXPERIMENT RESULTS

5.1. Efficiency and effectiveness results

In this experiment two main dependent variables were measured as indicators of efficiency and effectiveness:
• the length of time spent in executing the task;
• the number of incorrect answers in recognising the data structures, where the higher the number of incorrect answers, the lower the effectiveness.
For each subject, there are three 'timings' and three 'incorrect answers' results: one for each interaction method.
The averages, over the number of subjects, for the timings and the number of incorrect answers, and their significance, were calculated.

5.2. Efficiency results

Non-parametric Friedman's ANOVA tests were used to calculate the significance of the results because the test results were not distributed normally (condition for parametric tests). The results for efficiency and the relative statistical tests indicate that the Low interaction method is slower than the High interaction method ($p < 0.0167$). This means that exploring a sonification

moving the mouse freely up and down the timeline is more efficient than selecting a section of the sonification, selecting the speed at which to hear it and then pressing the play button. The efficiency of the Medium interaction method is not significantly different from the other two methods.
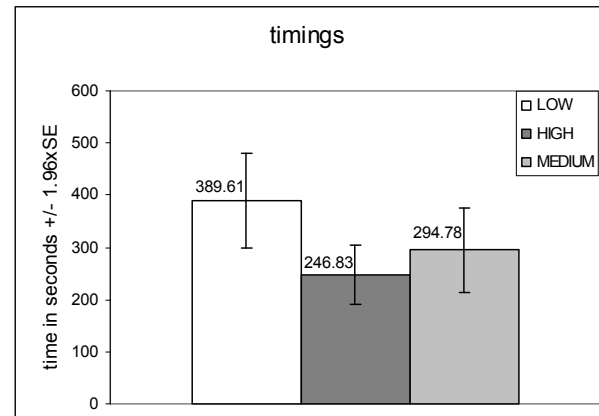

Figure 4: *Efficiency results*

5.3. Effectiveness results

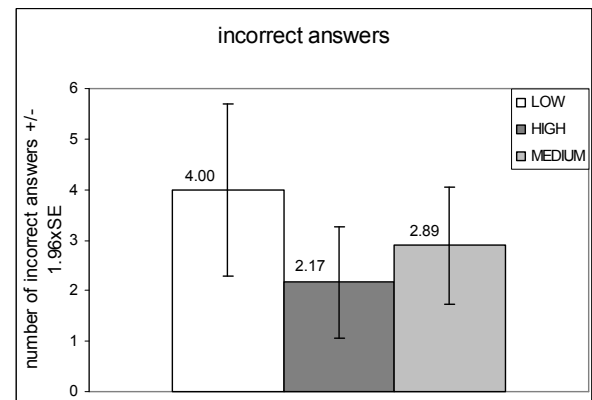In this experiment, no significant difference ($p > 0.5$) in effectiveness was found between the three conditions.

Figure 5: *Effectiveness results*

6. QUESTIONNAIRE RESULTS

In the questionnaire, the subjects were asked to score each interaction method for pleasantness, intuitiveness, clarity and quickness. At the end, they were also asked to express explicitly which interaction method they preferred. The results obtained from the questionnaire tell us how the different interaction methods were perceived subjectively. In particular, the results for quickness can tell us about the perceived efficiency of the interaction methods. The results for intuitiveness and clarity can tell us about the perceived effectiveness of the interaction methods.

6.1. Pleasantness

The Medium interaction method is found to be the most pleasant followed by the High interaction method and then the Low interaction method. This means that the subjects liked using the shuttle interface which allows changing speed and direction of playback quickly and has a set number of constant playback speeds. To use the mouse with a direct mapping between speed of movement of the mouse and playback speed was judged to be the second most pleasant method and to select sections of the sound, select the speed of playback and press play was considered the least pleasant method of interaction. All the comparisons are significant with $p < 0.0167$.
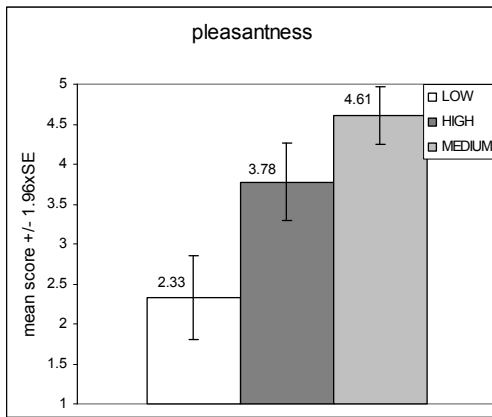

Figure 6: *Pleasantness results*

6.2. Intuitiveness and clarity: perceived effectiveness

The results for intuitiveness and clarity tell us that selecting a section of audio, choosing the playback speed and pressing play (Low interaction method) is considered significantly ($p < 0.0167$ for both comparisons) less intuitive and clear, than exploring the sonification using either the mouse or the shuttle interface. Differences between these last two methods are not significant.
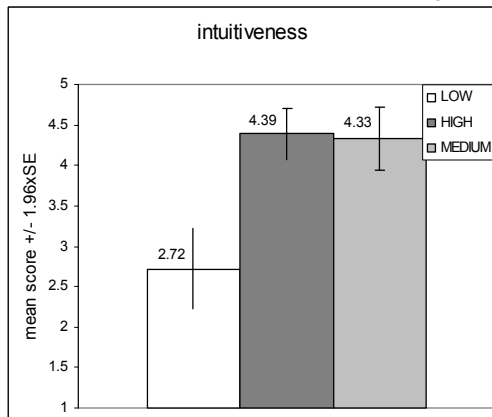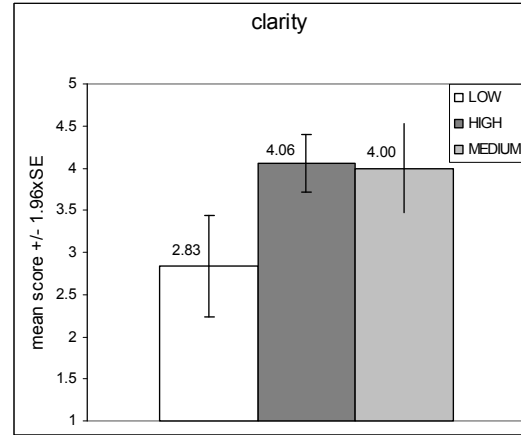

Figure 7: *Intuitiveness results*

Figure 8: *Clarity results*

6.3. Quickness: perceived efficiency

The Low interaction method is also perceived to be significantly slower than the other interaction modes ($p < 0.0167$ for both comparisons). There is no significant difference in quickness between the Medium and High interaction mode.
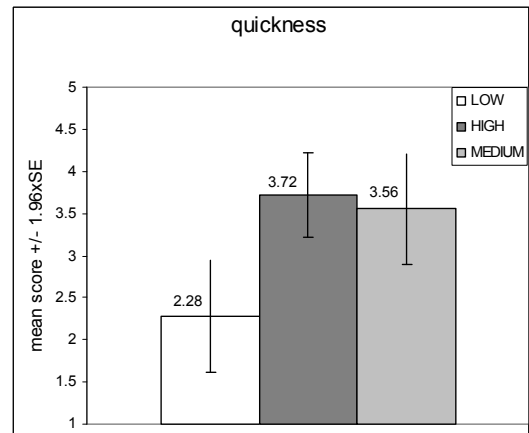

Figure 9: *Quickness results*

6.4. Preferred interaction method

The 18 test subjects were asked directly in the questionnaire which interaction method did they prefer:
• 4 out of 18 (22%) said the Low interaction method, i.e. selecting a section and a playback speed and pressing play;
• 6 out of 18 (33%) said the High interaction method, i.e. playing the sonification by moving the mouse backwards and forwards in the sound;
• 8 out of 18 (45%) said the Medium interaction method, i.e. playing the sonification using the shuttle interface.

6.5. Summary of overall questionnaire results

At the end of the test, each subject was asked to freely comment on the three proposed interaction methods. Below is a summary of the main points expressed in the comments. The numbers in brackets indicate the number of times a similar comment was made by different people.

	Advantages	Disadvantages
Low interaction method	* Better focus (3); * Better understanding of a specific region of data; * Familiar method to me; * Can be set very slow to avoid repeating the sound; * Constant speeds are comforting.	Very frustrating and hard to use
Medium interaction method	* More functionality (2); * One hand can be free (3); * Professional and appropriate; * Nice * It allows a constant playback speed; * Fine control at low speed; * Best control of playback and speed; * Intuitive and simpler; * No need to look at the screen;	* Difficult to match the audio with where you are in the sonification * Fixed play back speeds.
High interaction method	* The direct relationship between hand position and speed, and playback position and speed (3); * Total and quick control; * It allows for going straight to a certain point in the sound; * More intuitive and familiar; * Easier	None

Table 1: *Summary of subjects' comments*

The medium interaction method is judged to be the most pleasant and it is the preferred interaction method by the subjects. The High interaction method is the second best method and for intuitiveness, clarity and quickness it scores as well as the Medium interaction method. The Low interaction method is the worst in all the questionnaire results.

The Medium interaction method is considered appropriate, professional and nice. It has various functionalities (different fixed speeds, back and forward playback, "return to zero" button, fine control of slow speed, etc.) and works often hands free (i.e. hands can be used to do other things while listening to the display). It is not necessary to look at the screen area to know where one is in the data at any given time. This last aspect is at times considered confusing and can also become a disadvantage. The use of fixed playback speeds is both considered comforting, but at times restrictive.

The High interaction method is considered to give total control because there is a direct relationship between interface position and speed, and the playback position and speed. With this method, changes in position are immediate. The interface used is very familiar and results can be obtained very quickly.

6.6. Summary of overall results

The clearest result of this experiment is that the Low interaction method is considered the worst under all aspects.

The High and Medium interaction methods are considered better methods to use for navigating within a sonification data-space, and in particular the Medium interaction method is considered significantly more pleasant and was the preferred method of this group of subjects.

The main result from the objective measurements of efficiency and effectiveness is that the Low interaction method is slower, i.e. less efficient, than the other methods. The Low interaction method requires the user to perform many actions (select with the mouse, enter the chosen playback speed as a number, press play, etc.) before he/she can listen to the sonification. This procedure is time consuming making this method slower. The interaction methods were not significantly different in terms of effectiveness, i.e. the analysis of the sonification can be done equally well using all the methods.

The interaction methods reveal further clear differences when we look at how they are perceived by the user during the task.

The Medium and High interaction methods are perceived to be more efficient and effective than the Low interaction method as they score significantly higher for quickness, clarity and intuitiveness. Therefore, although from the objective measurements of effectiveness no difference was found between the methods, the subjects perceive the Low interaction method as significantly less effective than the other two.

Finally, subjects indicated in this experiment that, even if there are no particular objective differences between the Medium and High interaction methods, they prefer and find more pleasant the Medium interaction method which uses the shuttle interface. The reason for this preference could be the fact that the Medium interaction method provides, at the same time, very quick changes in playback speeds and direction, while allowing eyes and hand free moments in which the user can concentrate solely on the sound. The Medium interaction method does not require constant activity from users (the High interaction method does) allowing them to shift attention rapidly between different tasks, such as re-starting the sound, changing playback speed, listening to the sound and analysing the data.

7. CONCLUSIONS

From this experiment we can conclude that the addition of a relatively high level of interaction to a sonification display improves the efficiency of the analysis of the sonified data. From the point of view of the users, the addition of interaction highly improves the overall auditory display which is then perceived as more pleasant, clear, intuitive and quick to use.

Interesting ideas for further work would be to explore in detail the reasons for the subjects' preference of the Medium interaction method and develop and test a hybrid interface for

interactive sonification which groups the qualities of the mouse and the shuttle interface.

8.   REFERENCES

[1] S. Pauletto and A. Hunt, *"A Toolkit for Interactive Sonification"*, in *Proc. Of International Conference on Auditory Displays (ICAD)*, Sydney, Australia, 2004

[2] F. Winberg, and S. O. Hellström, "Qualitative aspects of auditory direct manipulation: a case study of the towers of Hanoi", *Proc. of the International Conference on Auditory Display*, Helsinki, Finland, 2001

[3] E. Brazil, M. Fernstrom, G. Tzanetakis, and P. R. Cook, "Enhancing sonic browsing using audio information retrieval", *Proc. of the International Conference on Auditory Display*, Kyoto, Japan, pp. 113–118, 2002

[4] T. Hermann, "Sonification for Exploratory Data Analysis" *PhD Thesis* in Faculty of Technology, Bielefeld University, Germany, 2002

[5] D. Di Filippo, and D. K. Pai, "Contact interaction with integrated audio and haptics", *Proc. of the International Conference on Auditory Display*, Atlanta, USA, 2000

[6] T. Beamish, K. Van Den Doel, K. MacLean, and S. Fels, "D'groove: A haptic turntable for digital audio control", *Proc. of the International Conference on Auditory Display*, Boston, USA, 2003

[7] J. Holmes, "Interacting with an information space using sound: accuracy and patterns", *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, 2005

[8] T. Stockman, G. Hind, and C. Frauenberger, "Interactive sonification of spreadsheets", *Proc. of the International Conference on Auditory Display*, Limerick, Ireland, 2005

[9] J. Kildal, and S. Brewster, "Providing a size-independent overview of non-visual tables", *Proc. International Conference on Auditory Display*, London, England, 2006

[10] T. Hermann and A. Hunt (eds.), "Special Issue on Interactive Sonification", *IEEE Multimedia*, April-June, Vol. 12, No. 2, 2005

[11] www.interactive-sonification.org

[12] Contour Design: http://www.contourdesign.com

[13] A. S. Bregman, *Auditory Scene Analysis*, Cambridge, Massachusetts: MIT Press, 1994