

SONIC BROWSING: AN AUDITORY TOOL FOR MULTIMEDIA ASSET MANAGEMENT

Mikael Fernström

Eoin Brazil

Interaction Design Centre
Department of Computer Science and Information
Systems,
University of Limerick, Ireland
Mikael.fernstrom@ul.ie

Interaction Design Centre
Department of Computer Science and Information
Systems,
University of Limerick, Ireland
eoin.brazil@ul.ie

ABSTRACT

In previous work, the *Sonic Browser* was used for browsing large data sets of music [1]. In this paper, we report results from an updated version of the Sonic Browser for managing general sound resources on personal computers. In particular, we have evaluated browsing of everyday sounds. The investigation was directed at comparing browsing single versus multiple stream audio. The problem of sound resource browsing for multimedia designers is the specific area of focus for our experiment. Finally, we conclude with current trends of our research for further improvement of the system.

1. INTRODUCTION

Browsing, in this context, is defined as “an exploratory, information seeking strategy that depends upon serendipity ... especially appropriate for ill-defined problems and for exploring new task domains” [2]. The amount of multimedia information and resources available on personal computers and via the Internet has grown exponentially over the past decade, hence there is a growing interest in multimedia data retrieval and management. Visual presentation of information is important and for visual resources there are a number of well-established tools for browsing, e.g. thumbnail-images. For sound resources there are very few equivalent tools. With most sound editor packages and system tools, only a single sound can be heard at a time, and quite often after having to click or double-click on a sound file icon.

In everyday listening one is often exposed to hundreds of different sounds simultaneously and is still able to pick out important parts of the auditory scene. On a computer, this can be compared to auditory icons [3]. With auditory icons, many different factors affect our ability to differentiate and select between the sound sources. Exploration mechanisms for this type of audio data exist, e.g. the Sonic Browser version for browsing large data sets of music [4]. Sound resources can be divided into two major categories, speech and non-speech sounds. The latter category can be further divided into those that deal with everyday sounds and those that deal with music [5]. These applications utilise our ability to switch our attention between different sounds in the auditory scene, making use of the “cocktail party effect” [6]. With multiple auditory streams it is interesting to note that people have different ability to differentiate between multiple sound sources and the number of concurrent sound sources. A metaphor for a user controllable function that makes it visible to the user is the application of an *aura* [1]. An *aura*, in this context, is a function that defines the

user’s range of perception in a domain. The *aura* is the receiver of information in the domain.

When working with multimedia authoring, image resources are today quite easy to locate due to browse and preview mechanisms in, for example, Adobe’s PhotoShop™. For audio clips, tools like Sonic Foundry’s SoundForge™ only has preview (or rather ‘prehear’) of a single sound at a time, for example in a *file open* dialogue. This makes it quite time consuming and difficult to locate desired sounds clips, especially since many sound effect CD/CDROMs use either numbers as file names or cryptic abbreviations. In some cases, long file names assist in giving hints about the contents of sound files, but this is rarely the case. In visual browsing, we are able to see several different small images or thumbnails at a time.

To browse for sound files, the *Sonic Browser* gives the user the possibility to listen to several sound files simultaneously and to navigate through a stereo-spatialised soundscape. Using the *Sonic Browser*, properties of the sonic objects can be mapped to arbitrary features of the visual display. File size can be mapped to size of visual symbols, sampling rates to colour, symbol shape to file type and horizontal and vertical location to date and time. Alternatively, the features of the visual interface can be mapped to represent the content of the sound files. The users can at any time change these arbitrary mappings, to suit their needs. When symbols are visually displayed like this, the user can move a cursor around in the display. The cursor is surrounded by an *aura*, i.e. the range of the user’s virtual auditory perception. All sonic objects within the *aura* will be playing simultaneously, panned out in a stereo-space around the cursor. The relative loudness is in inverse square relation to the center of the *aura*, the cursor.

1.1. Scope of this study

In this study we have compared how users navigate among sound files using Microsoft Windows Explorer (with single stream sound, invoked by at least one mouse click) and the *Sonic Browser* with multiple stream stereo-spatialised audio activated by *cursor/aura-over-icons*, representing sound files. We tried both short, cryptic file names and long descriptive file names, e.g. “dog barking”, “car starting”, “ball bouncing”. The aim was to provide fast and direct access to the sounds, so users easily can explore a number of sounds in parallel. With tight coupling between the visual and auditory information, users get a good spatial idea of what objects that are available and how to navigate between them.

2. TASKS

The scenario for this study is sound file management in multimedia authoring. Users were requested to locate and select sound files to match short descriptive sentences. The user's actions were captured on video, which was then analysed.

The datasets used in this experiment were drawn from two sources, a commercial sound effects CD¹ entitled "Spectacular Sound Effects - Trains, Planes, and Automobiles" and a Creative Labs Sound Effects CD². The first source contained 54 sounds and was used for familiarisation. The second source contained 244 sounds and was used in the test phase. The length of the sounds varied from 0.1 to 16 seconds in the test dataset.

3. USERS

Six postgraduate students were recruited to partake in this study. All the users reported having normal hearing and experience in use of the Windows Explorer and multimedia authoring.

The users were divided into two groups of three. The first group used MS Windows Explorer to browse the data sets for selection of sounds matching the sentences. The second group used the *Sonic Browser* to browse the same data sets, again based on the given sentences. In each specific task, the users were allowed to move the cursor around freely in the GUI trying to find target sounds. Overall, for the eleven auditory tasks, several interesting browsing behaviours were observed.

4. EXPERIMENT

This experiment was an exploratory experiment to further our understanding of how the design of interface elements and components of an interface when working with sound resources. We examined differences in the use of the *Sonic Browser* and Windows Explorer. In particular, we collected formative data relevant to the understanding of auditory browsing.

4.1. Experimental Design

One aim of our experiment was to ensure that the tasks were as close to possible to real world situations. This approach has been used in various related areas such as information retrieval [7] as *simulated work task situation* and in image retrieval by Jose et al [8] and by Rodden et al [9].

The Windows Explorer application is used for viewing file systems in Windows environments. It is a standard approach derived from the Apple Macintosh Hierarchical File System, which present the user with the file system as a folder/file hierarchy. There are several views available when using the Windows Explorer but in the experiment we used only a single view where files and folders in the currently selected view are listed as shown in Figure 1. The *Sonic Browser* presents the entire dataset using a visual star-field display with objects represented by colour, shape, size and location according to the particular objects attributes as shown in Figure 2.

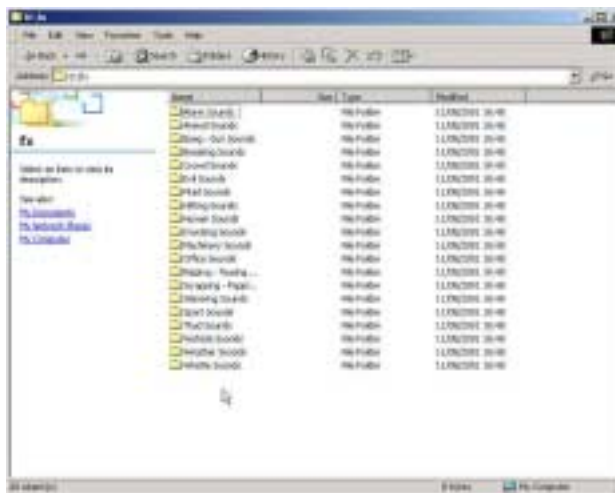


Figure 1: Windows Explorer

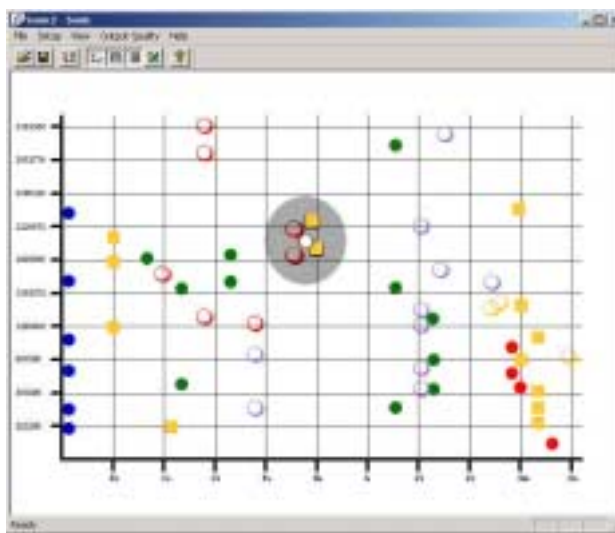


Figure 2: Sonic Browser

5. EXPERIMENTAL PROCEDURE

A normal high-end personal computer³ was used in the study. The computer was set up in an open plan office, which was similar to the normal work setting of the users. This configuration was used for both Windows Explorer and the *Sonic Browser*. The users' speech and actions were recorded on video.

The general method of evaluation used in this study is the *Thinking-Aloud* method [10], where participants are asked to voice their thoughts while trying to accomplish the tasks.

The participants were introduced to the aims of the study and the tools involved. They were then given some time to familiarise themselves with the tools. Then followed a test phase, and finally a debriefing. During the introduction, participants were shown the basic functions of the *Sonic Browser* and Windows Explorer.

At the beginning of the familiarisation, participants were given a list of practice tasks. The test phase was conducted in a similar way with a different sound data set and tasks. The debriefing was done immediately after the test. A questionnaire with semantic differentials was used for gathering the

participants' reactions to the features in each respective browser.

6. RESULTS

In the debriefing phase, a six point Likert scale questionnaire with five sets of semantic differentials was filled out by the participants who were asked to express their responses to the interfaces. (from 1 to 6, where 1 is "poor" and 6 is "excellent"). In Figure 3, the results of the questionnaire with cumulative participant responses displayed per question can be seen.

Questions one to four dealt with aesthetics, UI layout, layout interpretation and learnability. The results of these questions show that the users' prefer the Sonic Browser in terms of aesthetics and layout but only find a marginal improvement in learnability of this design. The users' only favoured the Windows Explorer interface over the Sonic Browser interface when the filename was known. This is seen in the result of question six. Questions seven and eight concerned how difficult it was to find a sound if the filename was unknown and how difficult would it be to find a group of related sounds. In question nine users' found it equally easy to play the last sound in either browser.

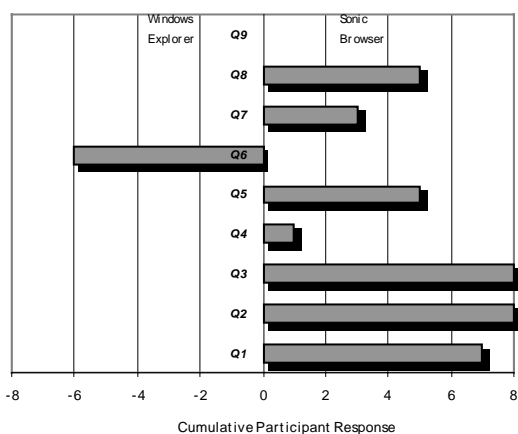


Figure 3: Results of questionnaire

The rich verbal protocol returned several interesting results during the experiment. It was found that a "Hierarchical Layout of folders was helpful for navigation", which also had a correlation to the "sounds can be easily spotted by file name" which relates to the long file names for files used in the data sets. This was not the case when using cryptic short file names. However, we found that by "extending the Aura over files, it allowed pan across files quickly" whereas with "Windows Explorer only has the ability to play one sound at a time".

Other issues were discovered in the debriefing and through user comments during testing, mostly relating to future improvements for the *Sonic Browser*. Several users' expressed a wish for both "a zoom facility" and "display by category" features. The first, to help navigate a clustered dataset and secondly to allow the user to choose only sounds that interested them and to temporarily hide the other sounds from the display. An issue was raised, which applied to both interfaces regarding "tagging" or "shading" of sound files. Several users' expressed a wish to mark "files of interest", and it was found that users' sometimes played the same sound several times and were looking for a mechanism to mark sounds which they had already examined and wished to "reference".

7. DISCUSSION

Commercial data sets with sound resources most often use cryptic filenames and no hierarchical organisation. Creating a hierarchy that allows for reasonable *view navigation* [11] requires substantial time and effort. A straight hierarchy view does improve performance, but can sometimes lead to ambiguity as users' categorization of sounds may differ. From our results, we can see that when a hierarchy with long descriptive file names is available, browsing will be faster unless categorization problems arise. In the case of categorization problems we have found that using an aura to browse makes it easier for the user to explore a sound data set.

8. FURTHER RESEARCH

In the case of the *Sonic Browser*, there are a number of design improvements that make sense in light of our experimental results.

Three *alphaslid*ers [12], each related to a specific attribute of a sound could be investigated. The first alphaslider would relate to what sound source, for example a car sound would be "car". The second would be associated with what action or event that is happening, for example, a sound of a car braking would be "braking". The final slider could exploit onomatopoeia [13], which is where sounds are described the way they sound, e.g. hammering could be "thunk-thunk". The sliders could be further enhanced by use of filter and zooming mechanism such as those found in the *LensBar* [14].

9. CONCLUSIONS

From the evaluation of the *Sonic Browser* and the Microsoft Windows Explorer interfaces we have shown that there are still further requirements for new mechanisms for browsing sound data sets. There are many new innovations from dynamic query research available for use in interactive systems. One such is the *Alphaslider*, which will be investigated as a technique for browsing large-scale sound dataset.

10. REFERENCES

- [1] J. M. Fernström and C. McNamara, "After Direct Manipulation - Direct Sonification," presented at ICAD '98, Glasgow, Scotland, 1998.
- [2] G. Marchionini, *Information Seeking in Electronic Environments*. New York, USA: The Press Syndicate of the University of Cambridge, 1995.
- [3] W. W. Gaver, "Using and Creating Auditory Icons," in *Auditory Display: Sonification, Audification and Auditory interfaces*, G. Kramer, Ed. Reading, MA, USA: Addison-Wesley Publishing Company, 1994, pp. 417-446.
- [4] J. M. Fernström and L. J. Bannon, "Explorations in Sonic Browsing," presented at BCS HCI '97, Bristol, UK, 1997.
- [5] K. Melih and R. Gonzalez, "Audio retrieval using perceptually based structures," presented at IEEE International Conference on Multimedia Computing and Systems, 1998.
- [6] B. Arons, "A Review of the Cocktail Party Effect," *Journal of the American Voice I/O Society*, vol. 12, pp. 35-50, 1992.

- [7] P. Borlund and P. Ingwersen, "The development of a method for the evaluation of interactive information retrieval systems," *Journal of Documentation*, vol. 53, pp. 225-250, 1997.
- [8] J. M. Jose, J. Furner, and D. J. Harper, "Spatial querying for image retrieval: a user-oriented evaluation," presented at ACM SIGIR conference on Research and development in information retrieval, Melbourne Australia, 1998.
- [9] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "Does organisation by similarity assist image browsing?," presented at SIG-CHI on Human factors in computing systems, Seattle, WA USA, 2001.
- [10] C. Lewis, "Using the "Thinking-aloud" Method in Cognitive Interface Design," IBM, Yorktown Heights, Research Report RC 9265, 1982.
- [11] G. W. Furnas, "Effective view navigation," presented at Human factors in computing systems, Atlanta, GA USA, 1997.
- [12] C. Ahlberg and B. Shneiderman, "The alphaslider: a compact and rapid selector," presented at Human factors in computing systems, Boston, United States, 1994.
- [13] S. Wake and T. Asahi, "Sound Retrieval with Intuitive Verbal Expressions," presented at ICAD '98, Glasgow, Scotland, 1998.
- [14] T. Masui, "LensBar - visualization for browsing and filtering large lists of data," presented at IEEE Symposium on Information Visualization, 1998.

¹ "Spectacular Sound Effect – Trains, Planes and Automobiles", copyright Thomas J. Valention Inc.

² "Creative Images, Sounds and Effects", copyright 1994 Creative Technology Ltd.

³ Intel Pentium III, 800MHz, 264 MB RAM, 19" display with 1280 x 1024 pixels at 85 hertz in 32-bit colour, Creative Labs SoundBlaster Live! Platinum sound card, loudspeakers Harman/Kardon HK19.5 speakers 2 X 12 at 10% THD, Microsoft Windows 2000 Professional v5.0.2195 Build 2195.