WHERE'S THAT SOUND? EXPLORING ARBITRARY USER CLASSIFICATIONS OF SOUNDS FOR AUDIO COLLECTION MANAGEMENT

Eoin Brazil, Mikael Fernström

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

ABSTRACT

Collections of sound and music of increasing size and diversity are used both by general personal computer users and multimedia designers. Browsing audio collections poses several challenges to the design of effective user interfaces. In this paper, we report results from a new 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 of audio resources with arbitrary classifications. 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 approach 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" [1]. 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. This is referred to as the data availability paradox by Woods [2], more and more data is available, but our ability to interpret what is available has not increased. In the context of this research, we are interested in facilitating human browsing behaviour focusing on the browsing of audio collections.

There has been much research into the area of musical information retrieval in the past few years as illustrated in review by Futrelle and Downie [3]. Music Information Retrieval has however focused mainly on the requirement to search databases of music using melodies as the base for queries. Limited work has been done regarding tools exploiting generalised audio in the areas of both searching and browsing. Generalised audio is all sound types and includes everything from speech, music, everyday sounds to other ambient sounds. Browsing sounds based on textual indexing alone can be difficult; for example, mp3.com classifies its songs into one of 215 different genres. Discovering new music amounts to downloading (and listening to!) an arbitrary number of songs, which match a particular text search. The results of this search could potentially number into the thousands. Users need a way to navigate and discover music files based on a variety of factors. Many applications offer only text-based searching by artist, song/album title, or music genre. In order to improve the user interaction with audio collections, novel user interfaces that allow simultaneous presentation of multiple sound clips for browsing are required.

Sound resources can be classified 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 [4]. The categories used for sound classification in the Sonic Browser are Filename, Environmental, Music, Speech, Onomatopoeia, Action/Event and Source. These categories have been influenced from previous work by Macaulay et al [5], by Gaver [6] and by Wold et al [7]. Macaulay et al classified elements of a soundscape based on sound type, information category and acoustical information. Gaver's hierarchical descriptions were used when the sound type was pre-classified as an everyday sound or an abstract sound. Wold et al proposed the classification of sound using simile, acoustical/perceptual features, subjective features and onomatopoeia. For recognition of sounds, while browsing, it is most often sufficient to hear only 500 ms to 2 seconds of the characteristic or significant part of a sound file [8]. Humans are remarkably good at genre classification as investigated by Perrot and Gjerdigen [9] where it is shown that humans can accurately predict a musical genre based on 250 milliseconds of audio. These recognition results were the basis along with research on the identification of sounds [10, 11] that suggest these classifications could assist in browsing of sounds, however these classifications are affected by the variance of individual perceptions and as such any form of classification of audio is highly subjective.

A new aspect of our research examines user-defined classifications of objects to allow for dynamic filtering of the browsing space. This dynamic filtering provides immediate feedback, which is tightly coupled to both the visual and auditory browsing space. Three types of filtering mechanisms are linked to the arbitrary user classifications of objects. The three mechanisms consist of text-based searching by category, dynamic sliders which are similar to AlphaSliders [12] and colour filtering. The shape and colour properties of an object are user definable and allow for assigning any property of the object in terms of shape or colour, e.g. an object representing a music file could have its musical key linked to the colour and its orchestration to the shape of the visual representation. In the Sonic Browser a central idea is to map properties of sound clips to aural and/or visual objects with properties that convey information about the sound clips and use these objects in order to create browsing spaces. The Sonic Browser is an application developed for accessing sounds or collections of sounds using sound spatialization and context-overview visualization techniques. The foundations of our design approach for the Sonic Browser are based on the principles of direct manipulation and interactive visualisation interfaces proposed by Shneiderman [13]. The three primary facets of this foundation are “overview first, zoom and filter, then detail on demand”. Exploiting this tight coupling a user can selectively attend to more information per unit time. Total information processing capacity is increased, thereby amplifying cognition. This guiding principle is explicitly stated as the principle of reducing the cost structure of information [14]. We avoid presenting a detailed introduction on the Sonic Browser[i] as this has been previously discussed by Brazil et al [15].

Many approaches have been developed to enable the searching or management of sound collections. Our work is one of the few attempts to present an interactive browsing system for digital audio collections that allows for multi-stream audio browsing and offers cross-platform compatibility.

1.1. Scope of this study

In this study we have compared how users navigate among sound files using the Sonic Browser with multiple stream audio activated by *cursor/aura-over-icons,* representing sound files. User navigation is supplemented by filtering mechanisms based on arbitrary classifications of the sounds. We used several types of sound files consisting of both real recordings of sounds and sound models (synthesised sounds). A sound model is a synthetic caricature of a sound, where the salient features of the sound have been identified and can be parametrically synthesised. The synthesized sounds in this study were created using Pure Data (PD), a software system for live musical and multimedia performances. PD files are also commonly referred to as patches or modules. The naming convention used for audio files was a short descriptive name of the sound. The aim was to provide fast and direct access to the sounds, so users can easily explore a number of sounds in parallel. With tight coupling between the visual and auditory information, users get a good spatial awareness of what sounds are present and how to navigate between them. The visual representations used in this study are 2-dimensional techniques, which include x-y plots (starfield display), Hyperbolic Trees and TouchGraphs [16-18]. A tight coupling also exists between the individual views where any changes in one particular view are mirrored to the other views.

2. Tasks

The scenario for this study is browsing within various 2-D visualisations for particular sounds matching a set criterion. As part of each task, the participants were asked to tag the sounds that they thought fulfilled the criteria of the task. The participants browsed the tightly coupled visual and auditory space manipulating it and applying specific filters to aid in the search for the sounds, meeting the desired task criteria. Participants were requested to listen to the sounds and to shift-click them to ‘tag’ a sound if they felt it matched the desired task criteria. In certain tasks they were asked to browse the collection after filtering it for specific properties. After selecting all the sounds, which they felt matched the criteria, the participant’s then saved a list of these tagged files via a menu option. The filtering of sounds within this study was done using property based filtering on arbitrary object properties.

The tasks included searching for a specific sound such as the `cry of a seagull' and to broader categories such as find all the sounds of `cats meowing'. In each specific task, the participants were allowed to move the cursor around freely in the GUI as well as change the current visualisation at will to compare relationships between sounds within different visualisations trying to find target sounds. Overall, for the eight auditory tasks, several interesting browsing behaviours were observed. At the end of each session, a questionnaire was presented to the participants in order to gain an insight into their feeling about the performed tasks.

The experimental data collection involved three techniques. First, data logging was collected by the application for the ‘tagging’ of objects in the 2-D space. Secondly, the user's actions were captured on video, which was then analysed. Thirdly, the participants filled out a post task questionnaire.

The dataset used in this experiment included 57 recorded sounds and 10 synthesized sounds, designed with PD-modules modelling impact interactions of two modal resonators [19], simplified so as to return only one mode. The recorded sounds used in this experiment were drawn from eight sources, seven commercial sound effects CD's and a local collection of ecological sounds. The length of the sounds varied from 0.1 to 50 seconds.

3. Users

Six postgraduate students were recruited to partake in this study. Five of the participants referred to have musical training in average of 6 years, with a minimum of 4 and a maximum of 12 years. One participant reported to have hearing problems with very low tones but the sounds used within the experiment weren’t affected by this condition. Two postgraduate students were recruited for the initial pilot stage of this study.

4. Experiment

This experiment was an exploratory experiment to further our understanding of sound collections, the requirements for their management and the design of interfaces and components for working with sound collections. We examined differences in the various browsing strategies and each participant’s perception of the sounds matching the various task criteria. In particular, we collected formative data relevant to the understanding of auditory browsing.

4.1. Experiment 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 [20] as *simulated work task situation* and in image retrieval by Jose et al [21] and by Rodden et al [22]. The Sonic Browser presents the entire dataset using one of three visualisations with objects represented by colour, shape and location according to the object's attributes as shown in Figure 1. The property filtering interface controls, used for several of the tasks are shown in the bottom left of Figure 1 and in greater detail in Figures 2 and 3. The HyperTree visualisation [17], a well-known focus + context technique, it is illustrated in the right hand side of Figure 1. The HyperTree uses a tree that is kept within the confines of a circular area on the screen and is based on a hyperbolic geometric transformation.

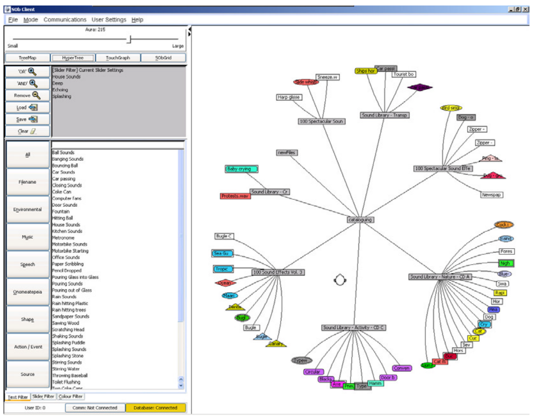

Figure 1: *The Sonic Browser HyperTree visualisation*

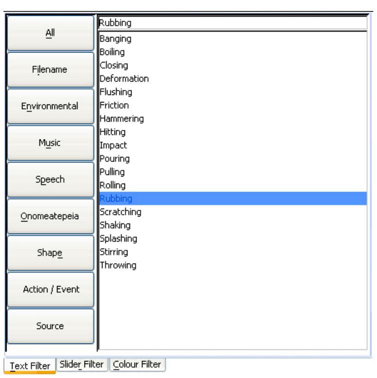

Figure 2: *The Sonic Browser text filter mechanism*

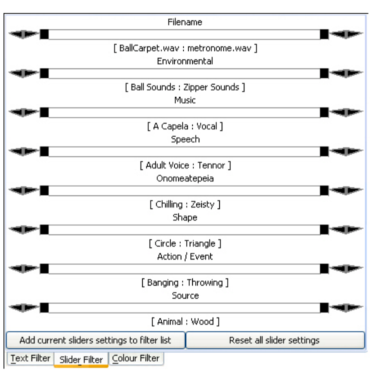

Figure 3: *The Sonic Browser dynamic slider mechanism*

5. Experimental procedure

A workstation Personal Computer[ii] was used in the study with the sounds being presented through stereo headphones. The users were allowed listen to the sounds as many times as they wanted and to 'tag' them if they felt the sounds matched the search criteria. The users' speech and actions were recorded on video.

The general method of evaluation used is this study is the *Thinking-Aloud* method [23], 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 application involved. They were then given some time to familiarise themselves with the application. After familiarisation was completed they were asked to complete the tasks after this a short questionnaire and debriefing was conducted. During the introduction, participants were shown the basic functions of the Sonic Browser.

The debriefing was done immediately after the tasks. A questionnaire with semantic differentials was used for gathering the participants' attitudes to the tasks and the application.

6. Results

The experiment measured qualitative differences in the various tasks between participants. The questionnaire was the main tool for analysis of results from the experiment. However, as the task and video data was collected during the experiment, some indication about the browsing strategies used by participants in the experiment was also obtained.

At the end of each session, as part of the participant debriefing, a questionnaire was presented to the participants in order to gain an insight into their opinion of the performed tasks and the Sonic Browser. The participants gave their responses to the interfaces and to the tasks filled out a seven point Likert scale questionnaire using the questions shown in Table 1 with six sets of semantic differentials. (From 0 to 6, where 0 is ``poor" and 6 is ``excellent"). In Figure 4, the results of the questionnaire with cumulative participant responses displayed per question can be seen.

Questions one, two and seven dealt with the aesthetics, interpretation and learnability of the Sonic Browser. The results of these questions show that the users' find the Sonic Browser easy to learn and use. Questions three to six dealt with the filtering mechanisms of the Sonic Browser. The results of these questions, confirmed by video analysis, illuminate several issues such as it is always easier to find a sound when you know its filename and that the filtering mechanisms were found to be easy to use. Question eight concerned a slight delay when playing an audio file with the Sonic Browser. The results show participants did not notice or find any difficulty with the delay and that it did not affect the participant's ability to complete the task. This allows us to say that while any delay in audio playback is still important, there is scope for a minor delay in a cataloguing scenario. Questions nine and ten deal with the realism and quality of the sounds, which were found to be excellent by the participants.

Q1	How easy to interpret and understand this design
Q2	How easy is it likely to be to learn to use this design
Q3	To find a particular sound, when you know its filename
Q4	To find a particular sound, when you don't know its filename
Q5	To find a set of particular sounds with a specific property from a category
Q6	To perform an AND or an OR query using the filtering mechanism
Q7	Overall, how would you rate the ease of use or usability of this design
Q8	Was there a noticeable play lag before a sound was played
Q9	How realistic did you find the sounds
Q10	How did you find the quality of the sounds

Table 1: *Survey questions*

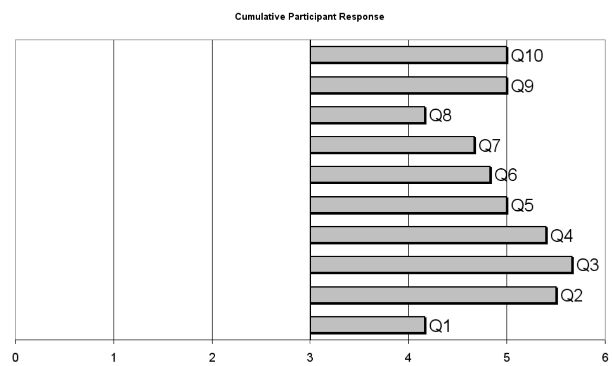

Figure 4: *Results of questionnaire*

The *Think-Aloud* method returned several interesting results during the experiment. One participant remarked they found a noticeable playback delay and stated that they ``expected sound to be instantaneous, not take ages". This was more likely related to the issues of silence at the start of the sound or a sound beginning at very low volume in the instance of the participant's comment. The HyperTree visualisation was preferred by half of the participants. Other issues were discovered in the debriefing and through user comments during testing, mostly relating to future improvements of the Sonic Browser. One suggested improvement includes adding more intelligent pattern matching to the filter property text lookahead field so that if 'Flushing toilet' is typed that it will also look for 'Toilet flushing'. Another improvement would be to add a ``Right click context menu with options for setting object properties and for group of objects properties", which would allow for faster setting of properties for both single objects as well as for groups of objects.

7. Discussion

An overview supplemented with property filtering does improve performance, but can sometimes lead to ambiguity as users' classifications of sounds may differ. From our results, we found that when a collection uses arbitrary classifications, such as descriptive file names, specific sound source, or sound event classifications, browsing will be faster unless classification problems arise. We also found that a hyperbolic layout for browsing makes it both easy and enjoyable for a participant to explore a sound data set.

The number of subjects who participated to the experiment was too small to obtain statistical significance; however, results suggest slightly better performance for using arbitrary classifications filtering in conjunction with browsing than for the browsing alone. In addition, we found that the experiment was useful for the gathering participants' comments and suggestions for future application functionality. Exploring sound collections is difficult, but it can be made easier through the use of dynamic filtering combined with both direct manipulation and direct sonification of the sounds in the collection.

8. Future research

After examining the results of our experiment, there are a number of issues that have been illuminated. One direction involves how to best to describe and present filter categories and the items within categories. Another direction of interest would be to explore alternative techniques for visualisation and object grouping. As participants found the hyperbolic layout engaging and enjoyable, another possibility for future work is the investigation of a hyperbolic three-dimensional space technique such as the H3 layout technique defined by Munzner [24]. The XML3D technique [25], a more recent technique by Munzner, is also being considered for investigation of its suitability. These techniques, in conjunction with arbitrary classification of sounds within a collection, might be examined as a possible method for browsing and managing large audio collections. A larger scale experiment examining the correlation between performance and participants' subjective measures of the tasks should be carried out. Another possible experiment would be to quantitatively compare different interfaces for sound browsing.

9. Conclusions

From the exploration of browsing sound collections we have shown that there are still further requirements for new mechanisms for browsing sound collections. The combination of dynamic query research with direct sonification offers new directions for research into interactive browsing systems.

10. Acknowledgements

This work was supported by the European Union under the European Commission's Future & Emergent Technologies collaborative R&D programme (Sob – The Sounding Object Project – no IST-2000-25287; http://www.soundobject.org). We would like to express our thanks to all the postgraduate students at the University of Limerick who participated in this experiment.

11. References

[1] G. Marchionini, *Information Seeking in Electronic Environments*. New York, USA: The Press Syndicate of the University of Cambridge, 1995.

[2] D. D. Woods, E. S. Patterson, E. M. Roth, and K. Christoffersen, "Can we ever escape from data overload?," presented at 43rd Annual Meeting on Human Factors and Ergonomics Society, Houston, Texas, 1999.

[3] J. Futrelle and J. S. Downie, "Interdisciplinary Communities and Research Issues in Music Information Retrieval," presented at 3rd International Conference on Music Information Retrieval, IRCAM-Centre Pompidou, Paris, France, 2002.

[4] K. Melih and R. Gonzalez, "Audio retrieval using perceptually based structures," presented at IEEE International Conference on Multimedia Computing and Systems, 1998.

[5] C. Macaulay and A. Crerar, "'Observing' the Workplace Soundscape: Ethnography and Auditory Interface Design," presented at ICAD '98, Glasgow, Scotland, 1998.

[6] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, pp. 1-29, 1993.

[7] E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," *IEEE Multimedia*, vol. 3, pp. 27-36, 1996.

[8] R. M. Warren, *Auditory Perception - A new Analysis and Synthesis*. Cambridge: Cambridge University Press, 1999.

[9] D. Perrot and R. O. Gjerdigen, "Scanning the dial: An exploration of factors in the identification of musical style," presented at Society for Music Perception and Cognition, 1999.

[10] J. A. Ballas, "Common factors in the identification of an assortment of brief everyday sounds," *J. of Experimental Psychology*, vol. 19, pp. 250-267, 1993.

[11] B. Gygi, "*Factors in the Identification of Environmental Sounds*", Ph.D., presented to the Department of Department of Psychology at Indiana University, 2001.

[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] B. Shneiderman, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 2 ed. Reading, MA, USA: Addison-Wesley, 1992.

[14] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Information visualization: Using vision to think.* San Francisco: Morgan-Kaufmann, 1999.

[15] E. Brazil, J. M. Fernström, G. Tzanetakis, and P. Cook, "Enhancing Sonic Browsing Using Audio Information Retrieval," presented at International Conference on Auditory Display ICAD-02, Kyoto, Japan, 2002.

[16] B. Shneiderman, "Tree visualization with tree-maps: A 2-d space-filling approach," *ACM Transactions on Computer Graphics*, vol. 11, pp. 92-99, 1992.

[17] J. Lamping, R. Rao, and P. Pirolli, "A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies," presented at CHI '95, Denver, CO, USA, 1995.

[18] A. Shapiro, "*TouchGraph Project*", Sourceforge.Net Open Source Repository, January 2002, Available at *http://sourceforge.net/projects/touchgraph/*.

[19] M. Rath, "Sound design around a real-time impact model", In *Models and Algorithms for Sounding Objects*, pp 32-45. SOb Project, 2002, Available at *http://www.soundobject.org*.

[20] 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.

[21] 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.

[22] 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.

[23] C. Lewis, "Using the "Thinking-aloud" Method in Cognitive Interface Design," IBM, Yorktown Heights, Research Report RC 9265, 1982.

[24] T. Munzner, "H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space," presented at IEEE Symposium on Information Visualization, Phoenix, AZ, 1997.

[25] T. Munzner, "*Interactive Visualization of Large Graphs and Networks*", Ph.D. Dissertation, presented to the Department of Computer Science at Stanford University, 2000.

---

[i] An installer for the application and instructions for its installation are available at http://www.soundobject.org as well as a user guide for the application.

[ii] AMD Dual Athlon MP, 1500MHz, 1024 MB RAM, 19" display with 1280 x 1024 pixels at 85 hertz in 32-bit colour, Creative Labs SoundBlaster Live! Platinum sound card, stereo headphones Sony MDR-CD280, Microsoft Windows 2000 Professional v5.0.2195 Service Pack 3.