Enhancing Music Browsing

Denis Lebel
Computational Acoustic Modeling Laboratory
McGill University - Music Technology Area
Montreal, QC, Canada
denis.lebel@mail.mcgill.ca

***Abstract*— This paper proposes a novel design approach for music browsing systems, which simultaneously presents multiple streams of audio content to the user. The browsing techniques used in existing systems are reversed, taking the audio content as the basis for browsing rather than the metadata. The proposed design is meant to provide a more efficient, transparent, and fun way to discover new music and browse through large collections.**

I. Introduction

Current music browsing systems are mainly text-based and exploit the standard hierarchical file system structure, combined with a predefined tagging classification mechanism. The browsing techniques involve lots of inefficient mouse clicking to traverse the hierarchy, and too much undesired audio content is unnecessarily and blindly loaded. Discovering new music is usually a time consuming and mostly trial-and-error process. The transparency of the systems is usually poor as their design assumes that the user understands the given classification and also that the content of the collection is properly classified.

The system described in this paper addresses those issues by reversing the current browsing techniques employed. Rather than starting with a textual representation of the songs of a collection, the system actually plays the audio content and complements the browsing experience with the metadata to provide more insights to the user and ease his navigation. Some similarity measure between elements of a collection is assumed to exist a priori, allowing for semi-automatic selection of audio content by the system.

The efficiency of the system is improved by exposing the user to multiple audio streams simultaneously, using the human ability to focus his attention on one sound source when exposed to several sources. This is commonly known as the Cocktail Party Effect (Arons 1992). The design is further enhanced by constructing streams using short audio excerpts rather than full-length songs, which increases and diversifies the exposure of the user to the content of the collection.

The transparency of the system is addressed noticeably by displaying the metadata of the content of the audio streams being played. The visual feedback then created allows the user to establish a mapping between the textual and musical information, thus simultaneously increasing his knowledge of the collection.

The proposed system is specifically targeting online music stores to enhance the browsing experience of music collections whose content is largely unknown by the user. The remainder of this paper describes the design of this hypothetical browsing system. Some perceptual tests are also proposed as possible directions for future work to evaluate the system design.

II. Related Work

The Sonic Browser (Brazil and Fernstrom 2003) represents one of the main attempts at enhancing music browsing. It uses multiple streams of audio spatialized in the listener space and features a two-dimensional starfield to represent the collection (Figure 1). Audio files are illustrated as dots with customizable attributes (e.g., shape, size, color, etc.). A shaded circular area is used to identify the audio files located in the listener space that are currently played.

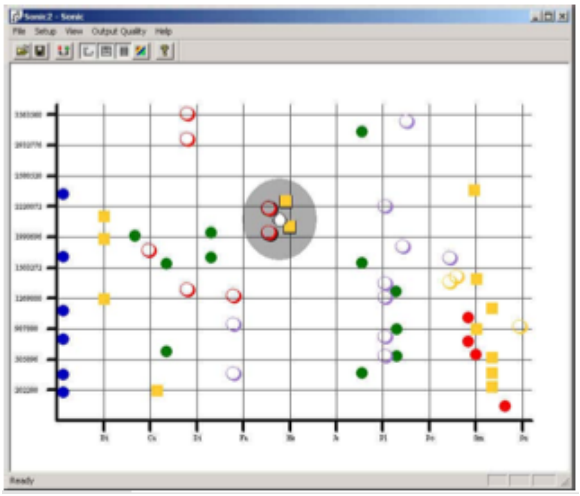

Fig. 1. Sonic Browser interface. The shaded circular area represents the audio files currently playing in the listener space. From (Brazil and Fernstrom 2003).

One drawback of the Sonic Browser is the clustering of sounds in the space, which creates confusion when too many similar sounds are played simultaneously. Moreover, even though representing collections using 2D maps seems like a trendy approach in recent literature (e.g., Goto and Goto 2005; Gulik et al. 2004), the information conveyed by these is pretty much limited to the concept of similarity. However, when searching for new content, the user ultimately needs to hear the music to determine whether he likes it or not. Thus, the user does not necessarily need to visualize the similarity characteristics of the entire collection.

III. Proposed System Design

Considering new music is usually discovered through listening, the browsing system described in this section exposes the user primarily to audio content rather than textual or visual information. Instead of presenting the complete song collection using its similarity characteristics, songs are dynamically selected by the system and played to the user who decides whether it suits his preferences.

Simultaneous Presentation

Since the amount of audio content that can be played is limited by the temporal dimension of sound, a more efficient way to present the audio content is required. Similarly to how the human vision system is able to quickly scan through a collection of images simply by looking simultaneously at all thumbnails, the human hearing system is able to focus its listening attention on a single sound source in a noisy environment. Thus, when playing several streams of audio in the listener space, the user should be able to selectively listen to what he is more interested in.

The spatial distance between the sources should be kept maximal to help the segregation of the sound. Some experiments with a basic prototype seemed to suggest that four sources worked reasonably well with sources playing fairly different audio content. Increasing this number tended to create confusion. Also, sources need to be equidistant from the listener and the loudness should be roughly equivalent for all sources to avoid masking others. Figure 2 shows an example of listener space in which four sources are evenly positioned around the listener.

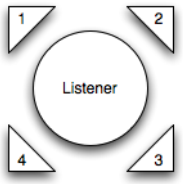

Fig. 2. Four sound sources evenly positioned around the listener.

Short-Term Exposition

It takes only a very short excerpt of audio to recognize a song. Similarly, a very limited exposure to a musical genre is often sufficient to decide whether it could fit our listening preferences. Thus, in addition to simultaneously playing multiple audio streams spatialized in the listener space, these could also be composed of short song excerpts queued up.

To further increase the exposure of the listener to the content of the music collection, the excerpts should be chosen as diversified as possible using some similarity measure. A dispatcher should then be used to fill the streams while keeping track of what has been played already (Figure 3).

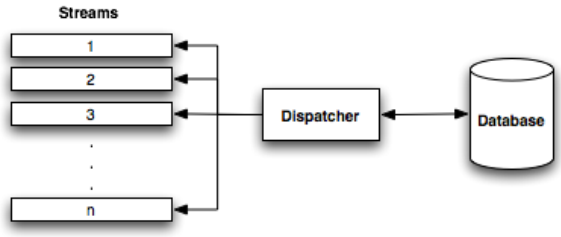

Fig. 3. Streams preparation using a dispatcher.

Asynchronous Events

Two consecutive excerpts in a same stream should be easily differentiable while transitioning smoothly from one to another (e.g., crossover). However, an offset should be introduced at the beginning of each stream such that transitions between segments of different streams happen asynchronously, as shown in Figure 4. The asynchronous events have the effect of introducing some temporal separation between the sources, which can act as a stimuli to help the listener alternate his focus on the different sources.

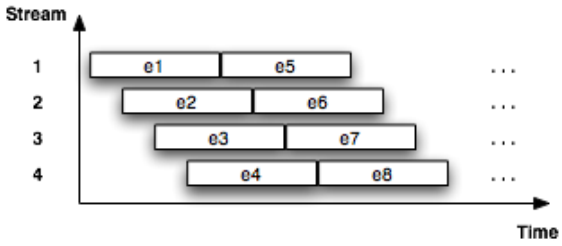

Fig. 4. Streams composed of short excerpts. Each stream has an initial offset to produce asynchronous transitions between segments of different streams.

Visual Feedback

To enhance the transparency of the system and its ease-of-use, it is important to provide visual feedback. In Figure 5, an example of a basic user interface is shown, in which album covers are displayed, mapping the playing position of the stream in the listener space. The artwork would change as the audio excerpts change, which would once again contribute to a better temporal separation of the sound sources. Metadata can also be displayed to provide information about the audio content.

Focus

Adding the possibility to focus on a specific source is essential. This can simply be done by modifying the loudness of an audio streams when the user focuses on one specific stream. With segmented streams, it causes the additional problem of having to loop the current segment to avoid transitioning with the next queued segment. Once the focus is released, the sound level returns to its normal and streaming continues as usual.

Selection

When the user selects a genre or a song he likes, the content of the streams is reorganized with song excerpts that are similar to the selection of the user. The dispatcher selects song excerpts with a higher similarity threshold, which means that the content of the streams will get less diversified. Selection could in fact occur as long as there is enough content left in the collection to fill the streams without looping (which should take quite a long time for large collections). The inverse operation is simply accomplished by lowering the similarity threshold value, thus allowing more diversified content in the streams.

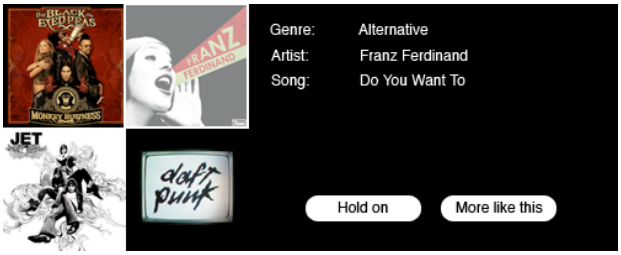

Fig. 5. An example of a basic user interface with visual feedback.

IV. Conclusion and Future Work

A novel system for browsing music collections was proposed as a more efficient and transparent solution for discovering new music. In this system, multiple audio streams, composed of short and diversified song segments, play simultaneously in the listener space while corresponding metadata information is displayed in the user interface. Selection of an excerpt effectively changes the diversity level of the content of the streams.

Several tests could be performed in regards to the efficiency of the proposed system compared to existing systems. Furthermore, perceptual tests are necessary to determine an appropriate length for the segments as well as a good time difference between the asynchronous transitions. More formal listening tests are also necessary to determine the ideal number of sound sources that can be used in such a system.

Acknowledgment

The author would like to thank Ichiro Fujinaga who initially proposed this very interesting project idea. Thanks also to everyone who took his *Music Information Acquisition, Preservation, and Retrieval* seminar this Winter for all the great discussions and suggestions.

References

[1] Arons, B. 1992. A review of the cocktail party effect. *Journal of the American Voice I/O Society* 12: 35-50.

[2] Brazil, E., and M. Fernstrom. 2003. Audio Information Browsing with the sonic browser. *Proceedings of the Conference on Coordinated and Multiple Views in Exploratory Visualization*. 26–32.

[3] Gulik, R., F. Vignoli, and H. Wetering. 2004. Mapping music in the palm of your hand, explore and discover your collection. *Proceedings of the International Conference on Music Information Retrieval*. 409–14.

[4] Goto, M., and T. Goto. 2005. Musicream: New music playback interface for streaming, striking, sorting, and recalling musical pieces. *Proceedings of the International Conference on Music Information Retrieval*. 404–11.

[5] Kobayashi, M., and C. Schmandt. 1997. Dynamic soundscape: Mapping time to space for audio browsing. *Proceedings of the Conference on Human Factors in Computing Systems*. 22–7.