INTERACTIVE CONTENT-AWARE MUSIC BROWSING USING THE RADIO DRUM

Jennifer Murdoch and George Tzanetakis

(gtzan,jmurdoch)@cs.uvic.ca
Department of Computer Science, University of Victoria, Canada

ABSTRACT

Portable digital music players are becoming pervasive and the size of personal digital music collections has been steadily increasing (5-10 thousand tracks are common today). The emerging area of Music Information Retrieval (MIR) deals with all aspects of managing, analyzing and organizing music in digital formats. The majority of work in MIR follows a search/retrieval paradigm. More recently, the importance of browsing as an interaction paradigm has been realized and several novel interfaces have been proposed. In this paper, we describe a tangible interface for content-aware browsing of music collections. The Radio Drum is a gestural interface based on capacitance sensors that can detect the x,y,z positions of two drum sticks in a 3D volume. We describe two possible mappings that can be be used for browsing music collections without relying on metadata. The first is an explicit mapping of tempo and beat strength, and the second is a music similarity space using audio feature extraction and a Self Organizing Map (SOM).

1. INTRODUCTION

The popularity of portable music players and digital music distribution have made large collections of music (thousands of songs) instantly accessible to everyday users. However, as any user with a large music collection can testify, browsing and searching the collection becomes increasingly harder as the collection grows. The identification of meaniningful relations between music recordings is complex for two primary reasons; the number of low-level and high-level music descriptors is seemingly endless, but in addition, the importance of each of these descriptors is coloured by the listener's own perception. Paradoxically, it is this variation in perception that both gives rise to the diversity of musical styles and tastes while driving the subjectivity and controversy within our attempts to organize them. Music Information Retrieval (MIR) is an emerging research area that explores how computers can be used to effectively interact and analyze large collections of music in digital format. Because music experience is subjective it is important to support personalization.

In recent years, there has been a trend towards automation of both collection organization, playlist generation and similarity retrieval. However most of the proposed approaches follow a conventional search/retrieval paradigm where there is a well-defined query or goal. In contrast browsing refers to the process of navigating through a collection of documents or musical pieces relying on serendipity rather than a specific target. Browsing is an inherently interactive activity requiring constant interplay between the user and the computer and fast response times. Another possible limitation of existing systems is that they are utilize a monitor/keyboard/mouse interface which can be cumbersome and limiting. The perception and appreciation of music is an interactive process which we feel is complimented best by an interactive approach to exploring music collections.

In order to motivate our approach we propose the following thought experiment: imagine that you want to browse a collection of five thousand songs for which you have no metadata or visual representation available. You are only allowed to interact with the system through your physical actions and your only feedback is through your ears. We claim that such a system would be an effective tool for interactive browsing of music collections. The addition of metadata and/or visualization will only add to the power of the system.

In this paper, we describe a prototype for such a non-visual interactive music browsing system. The system supports interactive, personalized browsing based on tactile control and continuous auditory feedback. More specifically we utilize an interactive gesture-sensing interface known as the Radio Drum and automatically create spaces for music exploration through automatic feature extraction combined with a Self-Organizing Map (SOM) for clustering and dimensionality reduction. Two mappings one based on rhythm and another based on music similarity are also described. The system utilizing the radio drum is rather expensive and targeted to expert users such as DJs. A lower cost version targeted to average users and utilizing the STC-1000 [1] interface by the Mercurial Innovations Group is also described.

This work was partly supported by a NSERC discovery grant, Memotrax Inc. and Mercurial Innovations Group.

[1] `http://www.thinkmig.com/stc1000.html`

2. RELATED WORK

One of the most inspiring interfaces for our work is *MusicBottles*, a tangible interface designed by Hiroshi Ishii and his team at the MIT Media Lab. In this work, bottles can be opened and closed to explore a music database of classical, jazz, and techno music [1]. This work combines a tangible interface with ideas from information retrieval. Similarly in our work we try to use existing interaction metaphors for music information retrieval tasks.

The idea of using a "TimbreSpace" for representing sounds and their relations was introduced by [2]. The Sonic Browser [3] introduced the idea of continuous auditory feedback to navigate collections of sounds. Initial work in Music Information Retrieval (MIR) concentrated on algorithmic devopment rather than interactive systems. However in recent years, there has been a steady increase in interfaces for MIR. Self-organizing maps have been in used in Islands of Music [4] as well as the Databionic visualization [5]. Another interesting interface is Musicream [6] which is a new music playback interface for streaming, sticking, sorting and recalling musical pieces. The idea of personalization is explored in [7] where a music retrieval system based on user-driven similarity is described. A tangible interface for browsing music collection using a table metaphore is described in [8]. These are representative examples and by no means an exhaustive list. The system described in this paper draws ideas from many of this systems but is differentiated by the combination of automatic mappings of music to space and tangible interaction.

3. SYSTEM OVERVIEW

In order to support the presentation of the system components, we start by informally describing the interaction of a user with the system. The system automatically extracts content-information from a large collection of music using MIR techniques. The music pieces are then mapped onto the surface of the Radio Drum taking into account the content and similarity characteristics of the music. The user controls two sound streams (one for each stick) in a similar fashion to a DJ with two turntables. By moving the sticks on the surface the user can interactively explore the music space and also control the mixing of the two streams. There is continuous sound playing and therefore no need for a play/select button. The only visual feedback is seeing the position of the sticks on the surface.

The main controller utilized in this work is the Radio Drum. In addition, the STC-100 which a lower cost controller has also been used. In order for the system to be content-aware, state-of-the-art MIR algorithms are used for automatic feature extraction. The resulting feature space is mapped to two dimensions using a Self-Organizing Map (SOM). In this section we describe the individual components of the system. In the next section we describe how these components are integrated in specific mappings for effective music browsing.

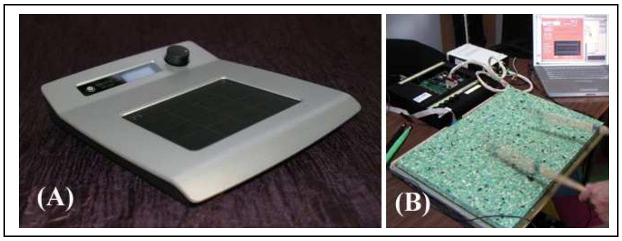

Fig. 1. (A) Mercurial STC 1000 and (B) Radio Drum

3.1. Radio Drum

The Radio Drum/Baton, shown in Figure 1 (B), even though not widely known, is a well established electronic music controller [9] used in academia and computer music performance. Built by Bob Boie and improved by Max Mathews, it has undergone a great deal of improvement in accuracy of tracking, while the user interaction has remained relatively constant. It consists of a detection surface, two drum sticks and a control box. The sticks don't have to be in contact with the surface for their position to be detected. The drum generates 6 separate analog signals that represent the current x, y, z position of two sticks. The radio tracking is based on measuring the electrical capacitance between the coil at the end of each stick and the array of receiving antennas on the drum (one for each corner). The analog signals are converted to MIDI (Musical Instrument Digital Interface) messages by a microprocessor and sent to a host computer. The sensing surface measures approximately 375mm X 600mm (15 X 24 inches). While the nature of the radio drum's design immediately lends itself to applications as a music synthesis interface, we present an application that uses the radio drum as a novel tangible interface for browsing and interacting with music collections.

As a lower cost alternative to the Radio Drum, the STC1000 controller has also been used. The Mercurial STC1000 (shown in Figure 1 (A)) [2] uses a network of fiberoptic sensors to detect pressure as position on a two-dimensional plane. It has been designed by the Mercurial Innovations Group. This device is a singe touch controller that directly outputs MIDI messages. The mapping to MIDI can be controlled by the user. The active pad area is 125mm X 100mm (5 X 4 inches).

3.2. Audio Feature Extraction

Automatic audio feature extraction is used in order to convert each music piece into a numerical representation that captures information about the content. In this work we utilize the feature set described in [10] for the purpose of automatic musical genre classification as well as tempo and beat strength estimation. The Marsyas software framework [3] is used for the feature extraction.

[2] http://www.thinkmig.com/stc1000.html
[3] http://marsyas.sourceforge.net

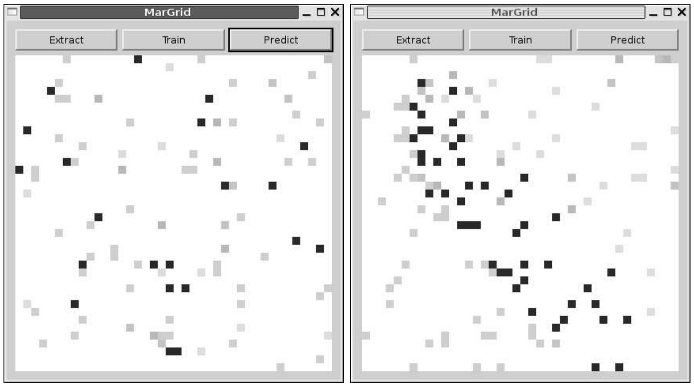

Fig. 2. Random placement of music pieces (Left), Self-Organized Map(Right). The colors(labels) are used only for displaying the structure of the map - the training is done in unsupervised fashion.

3.3. Self-Organizing Map

The Self-Organizing Map (SOM) is a type of neural network that clusters data of arbitrary dimensionality into two dimensions, thus facilitating both similarity quantization and visualization simultaneously. The SOM was first documented in 1982 by T. Kohonen , and since then, it has been applied to a wide variety of diverse clustering tasks. A good overview paper is [11]. In our application the SOM is used to map the high-dimensional audio feature space (30-dimensions) to two dimensions on the surface of the Radio Drum or the STC1000 pad.

The traditional SOM consists of a 2D grid of neural nodes each containing an n-dimensional vector, $\mathbf{x}(\mathbf{t})$ of data. The data associated with each node is initialized before training to small random values. During training, a series of n-dimensional vectors of sample data are added to the map. The *w*inning node of the map is found by computing the distance between the added training vector and each of the nodes in the SOM. This distance is calculated according to some pre-defined distance metric. Alternatively, a heuristic may be employed to search only part of the map for the most likely winning node, thus reducing the computational cost of adding each training sample. Once the winning node has been defined, this and those surrounding nodes reorganize their vector data to more closely resemble the added training sample. Two functions are thus introduced:

$\alpha(t)$ describes the size of a learning step, ie. to what degree the winning node's data, $\mathbf{x}(\mathbf{t})$, is altered to decrease the distane between the training sample and the node's current data. This value may change as training progress, and hence, it is a function of the time index t.

$neighbor\_val(d,t)$ describes the magnitude of the *s*phere-of-influence around the winning node that is affected by an added training sample along with the winner itself. The function should be defined for values of d, distance from the winning node, and t the time index for the added training sample. In our implement, $\alpha(t)$ is a linearly-decaying function with t. While some SOM applications define a discrete function for $neighbor\_val(d,t)$ such that a node is either within the winner's sphere-of-influence or is not, our implementation models the sphere of influence by a Gaussian function with un-normalized values with respect to integration area (to allow for values in the range [0,1]). The standard-deviation used in the Gaussian function is linearly-decreasing with t. This means that every node in the map is within the winner's sphere-of-influence to some degree as dictated by a Gaussian distribution, and this degree is dependent both upon the relative location of the node and the time index t. In a typical SOM application, both $\alpha(t)$ and $neighbor\_val(d,t)$ decrease with t. This facilitates convergence towards an "optimal" mapping, while at the same time allowing increasingly specific clusters of data to form within existing clusters.

Once a SOM has been trained, data may be added to the map simply by locating the node whose data is most similar to that of the presented sample, ie. the winner. The reorganization phase is ommitted when the SOM is not in training mode. Another interesting property of SOMs for our application is that they can be personalized by user initialization rather than random initialization.

Figure 2 shows two visualizations of mappings of musical pieces to space. Although these visualization are optional for the interaction with the system they are helpful for debugging and presentation. In these visualizations each piece of music is coded by a particular color. The colors are just used to convey visually the structure of the map. The training of the SOM is done in unsupervised fashion and only the automatically extracted feature vectors for each piece of music are provided. The picture on the left shows a random mapping of music pieces. The picture on the right shows the result of using the SOM for mapping. The dark cluster in the middle consists of pieces of rock music while the upper right corner contains classical music (light cyan in color) and the lower left corner contains hiphop and dance music (light red in color). This figure demontrates that the automatically extracted features capture information about the music content and that the SOM mapping preserve to a large extent that information.

4. INTERACTION

The main interaction scenario is browsing a collection of musical pieces rather than retrieval. The main idea is to provide a natural tangible interface for listening/browsing collections of music rather than the tedious playlist-play button model of existing music players. In all mappings the height from the surface is used to control the volume of each sound stream. For example lifting the sticks high up mutes both streams and bringing the stick to the surfaces maximizes the volume. That way arbitrary mixing of the two sound streams can be performed. In a DJ scenario one of the sticks would control the stream going into the headphones used to prepare the next song and the other stick would be used to control the currently playing song in a similar fashion to the use of two turntables.

Continuous sound playback is achieved using the idea of an aura, inspired by the Sonic Browser [3]. The aura is an imaginary sphere which is moved based on stick location. Any piece of music within the aura is mixed into the sound stream with volume inversely proportional to the distance from the stick location. That way pieces that are "closer" sound "closer" and vice versa. The size of the aura can be controlled. If there are many music pieces at a particular location the user can quickly scan through them by hiting the stick at that particular location. A lower cost variation of the system that with only one playback stream can be implemented using the STC1000. In that case pressure is used to control volume and the x,y position is used to navigate in space. The control is done with a finger in this case. A variety of possible mapping strategies can be used. In this paper we describe two mappings that we have found particularly effective.

4.1. Mappings

The first mapping is an explicit mapping based on rhythm where each axis is meaningful perceptually. Two perceptual dimensions characterizing rhythm are utilized: tempo and beat strength [12]. Tempo is mapped left-to-right with slow pieces left and fast pieces to the right and beat strength is mapped up-down with up being files with a strong sense of rhythm. The tempo and beat strength are extracted automatically. With this system a DJ can find songs at a particular tempo and beat stregth just by placing the finger or stick at the appropriate location.

The second mapping has no explicit axis interpretation but preserves topology and local similary neighborhoods by using the SOM for mapping the pieces of music to locations. The user has some control of the spatial arrangement of the map by providing proper initialization. For example if the user would like the map of Figure 2 to place the "black" pieces of rock music in a particular corner it could achieve it by initializing the map with a few rock pieces in that corner.

5. FUTURE WORK

There are many exciting directions for future work. Currently the mapping can not be edited while the user is interacting with the system. We are exploring the possibility of moving music pieces around using the sticks. This is especially challenging as it is hard to provide drag and drop or pick and drop interactions without visual feedback. Currently the map exists at one level. Multiple maps can be generated for different granularities but we would like to make zooming user controlled. We also plan to explore the use of a 5.1 audio playback system for placing the music pieces in particular spatial locations corresponding to their mapping. The system has been informally tested by users not involved with the development and their feedback has been very positive. A structured user study is planned for the future.

6. REFERENCES

[1] H. Ishii, "Bottles: A transparent interface as a tribute to Mark Weiser," *IEICE Transactions on Information and Systems*, vol. E87-D, no. 6, 2004.

[2] D. Wessel, "Low dimensional control of musical timbre," *Computer Music Joural*, vol. 3, no. 2, 1979.

[3] M. Fernstrom and E. Brazil, "Sonic Browsing: an auditory tool for multimedia asset management," in *Proc. Int. Conf. on Auditory Display (ICAD)*, Espoo, Finland, July 2001.

[4] A. Rauber, E. Pampalk, and D. Merkl, "Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structure of Music by Sound Similarity," in *Proc. Int. Conf. Music Information Retrieval (ISMIR)*, Paris, France, Oct. 2002, pp. 71–80.

[5] F. Morchen, A. Ultsch, M. Nocker, and S. Christian, "Databionic visualization of music collections acoording to perceptual distance," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[6] M. Goto and T. Goto, "Musicream:new music playback interface for streaming, sticking, sorting, and recalling musical pieces," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[7] F. Vignoli and S. Pauws, "A music retrieval system based on user driven similarity and its evaluation," in *Proc. Int. Conf. on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[8] I. Stavness, J. Gluck, L. Vilhan, and S. Fels, "The musictable: A map-based ubiquitous system for social interaction with a digital music collection," in *Int. Conf. on Entertainment Computing (ICEC05)*, Sept 2005, p. to appear.

[9] M. Mathews and W.A. Schloss, "The radio drum as a synthesizer controller," in *Proc. Int. Computer Music Conference (ICMC)*, Columbus, Ohio, 1989.

[10] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, July 2002.

[11] T. Kohonen, "The self-organizing map.," in *Proceedings of the IEEE*, 1990, vol. 78, pp. 1464–1480.

[12] G. Tzanetakis, G. Essl, and P. Cook, "Human perception and extraction of beat strength," in *Proc. Conf. on Digital Audio Effects*, 2002.