

Assistive Music Browsing using Self-Organizing Maps

George Tzanetakis
Department of Computer
Science
University of Victoria
gtzan@cs.uvic.ca

Manjinder Singh Benning
CANASSIST
University of Victoria
manjb@uvic.ca

Steven R. Ness
Department of Computer
Science
University of Victoria
sness@sness.net

Darren Minifie
Department of Computer
Science
University of Victoria
minofifa@uvic.ca

Nigel Livingston
CANASSIST
University of Victoria
njl@uvic.ca

ABSTRACT

Music listening is an important activity for many people. Advances in technology have made possible the creation of music collections with thousands of songs in portable music players. Navigating these large music collections is challenging especially for users with vision and/or motion disabilities. In this paper we describe our current efforts to build effective music browsing interfaces for people with disabilities. The foundation of our approach is the automatic extraction of features for describing musical content and the use of self-organizing maps to create two-dimensional representations of music collections. The ultimate goal is effective browsing without using any meta-data. We also describe different control interfaces to the system: a regular desktop application, an iPhone implementation, an eye tracker, and a smart room interface based on Wii-mote tracking.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
J.3 [Computer Applications]: Health

General Terms

Human Factors

Keywords

sound analysis, music information retrieval, content-aware user interfaces, assistive technologies, eye tracking

1. INTRODUCTION

Listening to music is a common everyday activity for many people. Recently the widespread availability of portable digital music players and digital music distribution have made

possible to carry around increasingly large personal music collections (currently 5-10 thousand tracks is a common size). As these collections keep getting larger navigating them becomes increasingly complex. Portable players and music track management software typically allow users to select artists, genres or individual tracks essentially by browsing long lists of text. This mode of interaction although adequate for small music collections becomes increasingly problematic as the collections become larger. The emerging area of Music Information Retrieval (MIR) [8] deals with all aspects of managing, analyzing and organizing music in digital formats. In the past ten years several MIR algorithms and user interfaces have been proposed that can assist with the browsing and navigation of large music collections.

Music browsing can be especially challenging for people with special needs. For example, finding a particular artist out of a long list of text using a scroll-wheel can be very difficult or impossible for a user with motor disabilities. Similarly reading text on a screen is not directly possible for a blind user. Various types of assistive technologies such as single-switch access, voice recognition and text synthesis which can be used for general computer usage, can also be applied in the case of music browsing. At the same time we believe that music browsing and listening are activities with unique and distinct characteristics that motivate the creation of domain specific software and control interfaces. Such tools have the potential to provide a much more effective and enjoyable experience for people with special needs. In this paper we report on our efforts to design such software and control interfaces for assistive music browsing.

The foundation of the system is a content-aware grid surface representation of a music collection that is calculated using Self-Organized Maps [6] over automatically extracted features that describe musical content. We have experimented with several control interfaces ranging from simple button navigation to location sensors based on the Wii-mote and eye tracking. This flexibility is crucial as there is a broad range of visual and motor disabilities that could potentially benefit from the proposed system. Each individual user with special needs has unique accessibility constraints so using the different control interfaces allows us to personalize the basic system. This personalization as well as the time availability of users makes it harder to conduct detailed quantitative user evaluations common in human-computer

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. PETRA'09, June 09-13, 2009, Corfu, Greece.
Copyright 2009 ACM ISBN 978-1-60558-409-6...\$5.00.

interaction. Instead we follow a more qualitative approach based on questionnaires and participatory design for guidance.

The context of this work has been *CanAssist*¹ which is a non-profit organization at the University of Victoria that specializes in the research and development of assistive technology as well as the development of community based initiatives to make life easier and more rewarding for those with disabilities. The technical group consists of professors, full-time engineers, graduate students on both the Masters and Phd level, and short-term student internships from a broad range of disciplines including electrical and mechanical engineering, computer science, music and fine arts, psychology, kinesiology, and neuroscience and medicine.

2. RELATED WORK

The two main research areas that have informed our work are assistive technologies and music information retrieval especially content-aware user interfaces for music browsing. There is a growing interest in human-centric computing which is a field where computers and devices are designed to adapt to the user need and preferences. Assistive technologies hold the potential to make life easier and more rewarding for people with special needs. Representative examples that have influenced our work include: smart environments for health care monitoring applications [3, 2], visual tracking of body features for computer access for people with severe disabilities [1], and non-visual interfaces for blind users [9].

Initial work in Music Information Retrieval (MIR) concentrated on algorithmic development rather than interactive systems. However in recent years, there has been a steady increase in the number of interfaces available for MIR. Self-organizing maps have been used in Islands of Music [?] as well as the Databionic visualization [?]. Another interesting interface is Musicream [?] which is a new music playback interface for streaming, sticking, sorting and recalling musical pieces. The idea of personalization is explored in [?] where a music retrieval system based on user-driven similarity is described. A tangible interface for browsing music collection using a table metaphor is described in [?]. Another tangible interface is *MusicBottles*, 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 [?]. Music collection browsing on mobile devices has also been explored [5]. An earlier version of the proposed SOM-based music browsing system with focus on DJ interaction using a Radio Drum control interface can be found in Murdoch [7]. The system described in this paper draws ideas from these systems but is differentiated by the diversity of control interfaces and the focus on accessibility.

3. SYSTEM OVERVIEW

The foundation of the proposed assistive music browsing system is a content-aware visual representation of a music collection. Automatically extracted audio features based on state-of-the-art audio signal processing techniques are used to represent musical content. A Self-Organizing Map (SOM) is used to map these high-dimensional audio feature vectors to 2D coordinates in a discrete rectangular grid. More details are provided in the following subsections.

¹<http://www.canassist.ca>

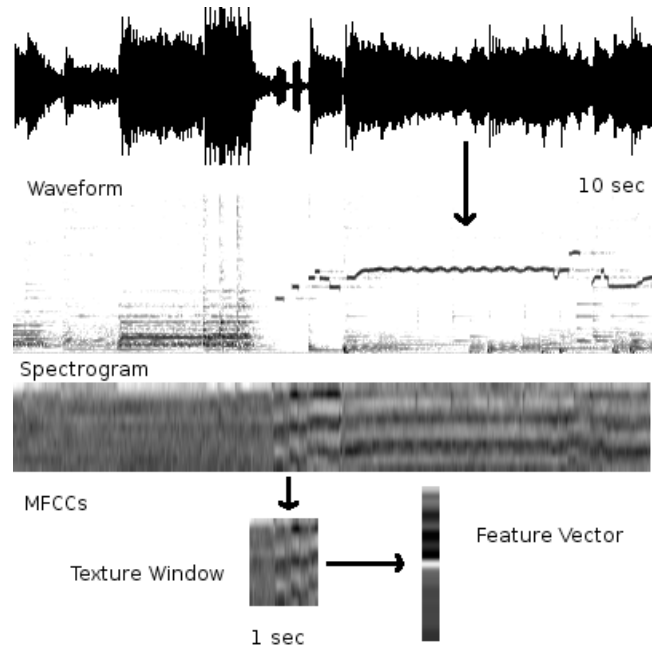


Figure 1: Feature extraction and texture window

3.1 Music Content Feature Extraction

The goal of this stage is to represent each song in a music collection as a single vector of features that characterize musical content. Using suitable features, songs that “sound” similar should have vectors that are “close” in the high dimensional feature space. The features used are the Spectral Centroid, Rolloff, Flux and the Mel-Frequency Cepstral Coefficients (MFCC). To capture the feature dynamics we compute a running mean and standard deviation over the past M frames:

$$m\Phi(t) = \text{mean}[\Phi(t - M + 1), \dots, \Phi(t)] \quad (1)$$

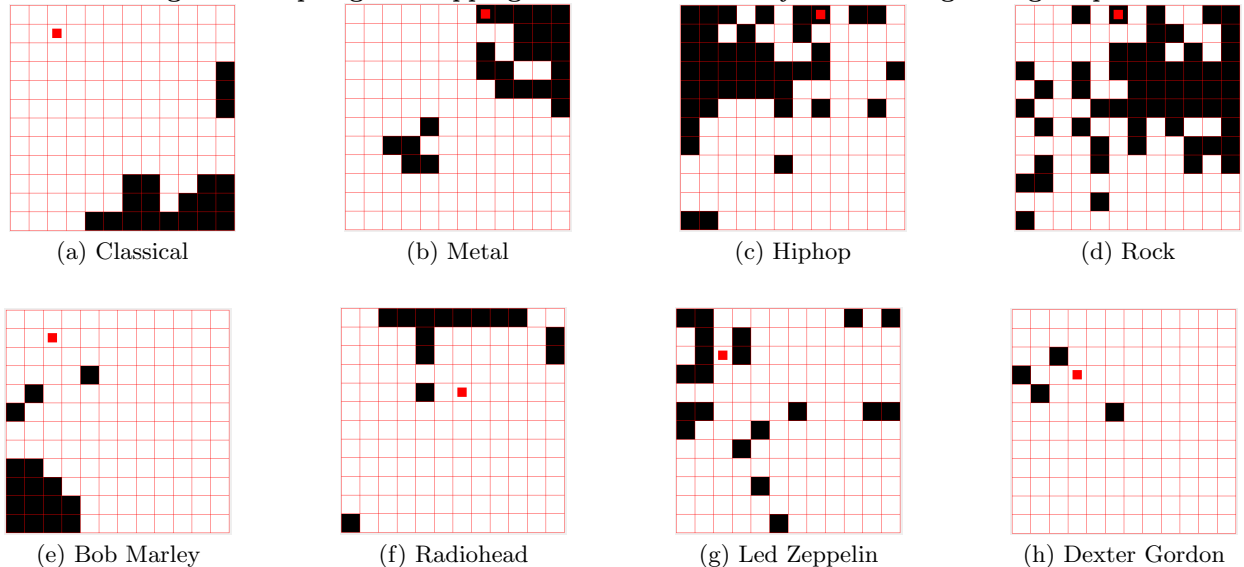
$$s\Phi(t) = \text{std}[\Phi(t - M + 1), \dots, \Phi(t)] \quad (2)$$

where $\Phi(t)$ is the original feature vector. Notice that the dynamics features are computed at the same rate as the original feature vector but depend on the past M frames (40 in our case corresponding to approximately a so called “texture window” of 1 second). This results in a feature vector of 32 dimensions at the same rate as the original 16-dimensional feature vector. This process is illustrated in Figure 1. The sequence of feature vectors is collapsed into a single feature vector representing the entire audio clip by taking again the mean and standard deviation across the 30 seconds (of the sequence of dynamics features) resulting in the final 64-dimensional feature vector per audio clip. A more detailed description of the features and their motivation can be found in Tzanetakis and Cook [11]. For the calculation of the self-organizing map described in the next section all features are normalized so that the minimum of each feature across the music collection is 0 and the maximum value is 1. This feature set has shown state-of-the-art performance in audio retrieval and classification tasks.²

3.2 Self-Organizing Maps

²<http://www.music-ir.org/mirex/2008>

Figure 2: Topological mapping of musical content by the Self-Organizing Map



The self-organizing map (SOM) is a type of neural network used to map a high dimensional input feature space to a lower dimensional representation while preserving the topology of the high dimensional feature space. This facilitates 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 [?]. In our system the SOM is used to map the audio features (64-dimensions) to two discrete coordinates on a rectangular grid.

The traditional SOM consists of a 2D grid of neural nodes each containing an n -dimensional vector, $\mathbf{x}(t)$ of data. The goal of learning in the SOM is to cause different neighbouring parts of the network to respond similarly to certain input patterns. This is partly motivated by how visual, auditory and other sensory information is handled in separate parts of the cerebral cortex in the human brain.

The network must be fed a large number of example vectors that represent, as closely as possible, the kinds of vectors expected during mapping. The examples are usually applied several times. The data associated with each node is initialized to small random values before training. During training, a series of n -dimensional vectors of sample data are added to the map. The “winning” node of the map known as the *best matching unit* (BMU) 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 which in our case is the standard Euclidean distance on the normalized feature vectors.

Once the winning node has been defined, it and its surrounding nodes reorganize their vector data to more closely resemble the added training sample. The training utilizes competitive learning. The weights of the BMU and neurons close to it in the SOM lattice are adjusted towards the input vector. The magnitude of the change decreases with time and with distance from the BMU.

The update formula for a neuron with representative vec-

tor $\mathbf{N}(t)$ can be written as follows:

$$\mathbf{N}(t+1) = \mathbf{N}(t) + \Theta(v, t)\alpha(t)(\mathbf{x}(t) - \mathbf{N}(t)) \quad (3)$$

where $\alpha(t)$ is a monotonically decreasing learning coefficient and $\mathbf{x}(t)$ is the input vector. The neighborhood function $\Theta(v, t)$ depends on the lattice distance between the BMU and neuron v . We utilize a Gaussian neighborhood function that shrinks over time. The time-varying learning rate and neighborhood function allow the SOM to gradually converge and form clusters at different granularities. In our implementation, $\alpha(t)$ is a linearly-decaying function with t . 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 omitted when the SOM is not in the training mode. Another interesting property of SOMs for our application is that they can be personalized by user initialization rather than random initialization.

Figure 3.1 illustrates the ability of the extracted musical content-features and the SOM to represent musical content. The top subfigures (a), (b), (c) and (d) show how different musical genres are mapped to different regions of the SOM grid (the black squares are the ones containing one or more songs from each specific genre). As can be seen Classical, Heavy Metal and HipHop are well-localized and distinct whereas Rock is more spread out reflecting its wide diversity. The SOM is trained on a collection of 1000 songs spanning 10 genres. The bottom subfigures (e), (f), (g), (h) show how different artists are mapped to different regions of the SOM grid. The SOM in this case is trained on a diverse personal collection of 3000 songs spanning many artists and genres. It is important to note that in all these cases the only information used is the automatically analyzed actual audio signal and the locations of the genres/artists are emergent properties of the SOM. If greater grid granularity is desired by a user, this system can also be used in parallel with an automatic genre classification system to reduce the total number of songs.

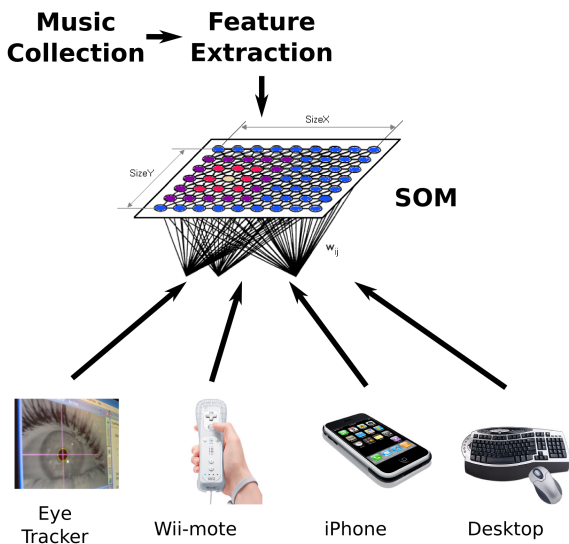


Figure 3: An overview of the system

4. CONTROL INTERFACES

The goal of our system is to provide flexible, personalized ways of music browsing for users with special needs. To this effect we have explored several different control interfaces to the same underlying content-aware representation for music collections. Figure 3 shows the overall architecture of our approach. The Marsyas³ audio processing software framework has been used for the audio feature extraction, calculation of the SOM, the graphical user interface and handling of controller data [10]. The basic system configuration utilizes the standard display/keyboard/mouse interaction familiar to most computer users. Although familiar this desktop metaphor is not suitable in many situations especially for people with visual or physical impairments. In the following subsections we describe alternative control interfaces that can be used for that purpose.

4.1 Mobile phone touch screen

New generations of mobile devices from Apple, Blackberry, Nokia and HTC provide a touch-based user interface instead of the traditional physical keyboard interface. While still in its infancy, touch technology has already refined the way users interact with a mobile device. Apple has patented multi-touch technology, whereby the user may use multiple fingers to control a device with complex gestures. Yet to be addressed is the decoupling of touch input from visual onscreen objects. As an example, the iPhone presents audio playback controls to the user as a series of icons. The user can then touch the icon to perform the associated action. This approach is inhibitive to the visually disabled because they may not be able to see the graphical controls in the first place. A better approach is to rely on spatial touch, where the user can feel areas of the touch display and associate them with the various controls.

Many portable devices include accelerometers that can

³<http://marsyas.sourceforge.net>

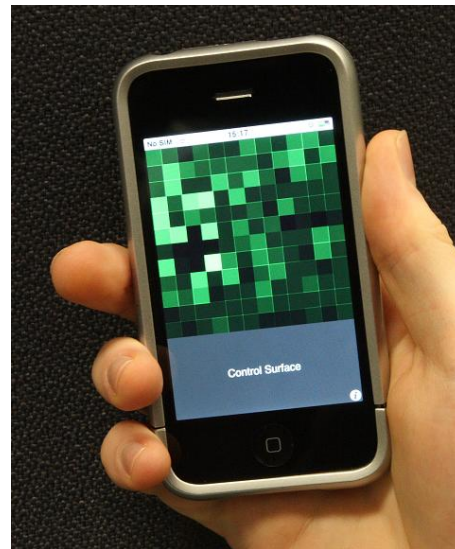


Figure 4: iPhone control interface

sense movement of the device in three dimensions. Many games on the iPhone take advantage of this by allowing the user to control the game through tilting, rotating and shaking the device. Similar user actions can be defined to control various aspects of music navigation and playback. An important advantage of using such mobile devices compared to custom-made controllers is that they can integrate seamlessly into the users day to day routine.

We have developed a music browsing prototype using the iPhone mobile device. The main window for the music browser is a grid that covers most of the screen. Each square on the grid represents a SOM node that holds a group of songs. When the user touches a square on the grid, a random song from that node begins to play. As the user moves her finger across the various squares, songs from each corresponding node cross-fade with each other to help her navigate the music collection by hearing how the songs in each grid location are changing. By laying out a music collection in this spatial fashion, navigation of the collection with only the knowledge of a few reference points is needed. For example, if it is known that Rock music is in the upper left corner, and Jazz music is in the lower left corner, by dragging from top to bottom along the left edge of the grid, Rock music will slowly transition into Jazz music. The use of multi-touch (two or more fingers) may also be used to control playback. Swiping the surface with two fingers in the right direction skips to another song in the same node while swiping left plays the previous song in that node.

We believe that user input via touch, rather than a traditional mouse or keypad, raises the level of usability and interaction between the user and their music collection. The large surface allows us to provide a tactile and spacial sensation, more appropriate for accessibility than a visual based, point-and-click interface. Figure 4 shows the interface running on the iPhone.

4.2 Location tracking using the Wii-mote

Recently Nintendo released a new game console with a novel controller called the Wii remote control, or Wii-mote. The Wii-mote has buttons like a standard game controller,

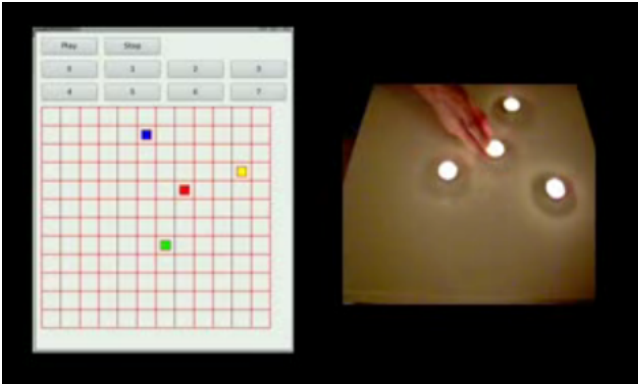


Figure 5: *Wii-mote interface*

but also has a 3D accelerometer and an infrared sensor. The combination of these two new input modalities has allowed game developers to create engaging new gaming experiences.

For this project we are using the Wii-mote in two different modalities. In the first of these, we use the infrared sensor on the Wii-mote to track a single infrared source above the display device. We then map the location where the user is pointing the Wii-mote to a location on the Self-Organizing Map containing the music collection. We also map the accelerometer of the Wii-mote to speed up and slow down the music and to apply different effects to the music. Buttons on the Wii-mote allow the user to pause playback and change the volume of the music track.

In the second modality, we again use the infrared sensor on the Wii-mote, but instead of having the Wii-mote move while the infrared source is stationary, we keep the Wii-mote in one position and move the infrared light sources. Specifically, we mount the Wii-mote on the ceiling and attach infrared light sources, in this case simple incandescent bulbs, to wheel chairs. Up to four infrared sources can be independently tracked by one Wii-mote, which means up to four motion impaired persons can be tracked as they move around a space. The infrared light sources. A simple example of this is shown in Figure 5 where the infrared sources are tea lights that can be moved on a surface at which the Wii-mote is pointing. The positions of the lights are tracked in real time and their data is sent to the self-organizing map interface. Although there are more advanced position tracking systems available, the Wii-mote has the advantage that it is a ubiquitous, low-cost device, and allows for quick prototyping of interaction scenarios.

In this case, instead of full songs being mapped in each grid coordinate, individual instruments playing different parts of the same song are mapped to each grid coordinate. For example in our canonical test case of Brazilian Samba music, one grid coordinate would contain the rhythm of the surdo drum, and other would have agogo bells, tambourim, shaker and cowbell. As the participants move around the map, each one activates a different rhythm. Together they can co-creatively generate music.

This system is particularly well suited for participants without fine motor control skills. The only requirement for this system is that they can locomote around in a room. We envision that this system could be used by a wide variety of users, including handicapped children and adults, and could

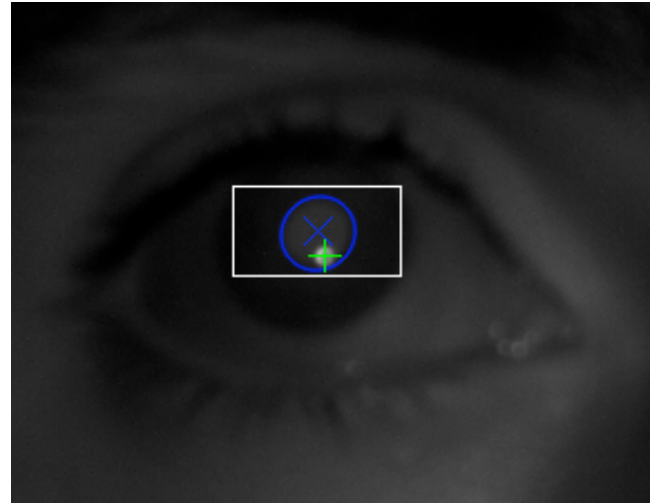


Figure 6: *Eye snapshot with gaze detection*

also be employed in a Senior Citizens home as a way for multiple participants to create music together.

4.3 Eye tracker

The CanAssist eye tracking system performs non-intrusive, accurate point of gaze (POG) estimation independent of head movements. The system uses the bright-eye, corneal reflection, image differencing approach with infrared sources to accurately track the POG of a user [4]. To achieve the level of accuracy required for the application of music browsing, a calibration is required. The user is instructed to follow a dot that moves and waits at four locations on the computer screen. The dot will not advance to the next location unless the user's sampled gaze points are within a certain standard deviation. This is to ensure that a satisfactory calibration is performed.

Figure 6 shows the pupil of the eye being tracked with a cross marking the centre. The plus sign marks the centre of the reflected white corneal glint. The square around the pupil show the region of interest that is searched to find the corneal glint. A position vector defined between cross and the plus sign is used as input to the mapping algorithm to estimate the point of gaze.

5. QUALITATIVE USER EVALUATION

Evaluation of browsing interfaces is challenging and becomes even more so in the case of assistive music browsing. The fact that each person with special needs has distinct constraints and abilities makes it impossible to calculate statistical results over multiple users. Therefore rather than following a quantitative evaluation methodology we have conducted a qualitative questionnaire-based evaluation and included disabled users in the design process. We have been fortunate to receive useful feedback from two users with special needs: one with cerebral palsy who used his lips to press keyboard keys for navigation and one with limited vision who relied mostly on touch and hearing. In addition an able-bodied user was also asked for feedback.

There were 5 control interface configurations although not all of them were tested by all users: desktop computer/mouse, desktop computer/keyboard, eye tracker, Wii-mote and iPhone.

The users were asked to perform two sets of tasks repeated over two phases followed by a questionnaire. For all the tasks a collection of 1000 songs was used. Task 1 requested users to find a song from a specific genre (Classical, Heavy Metal, Hip-Hop, and Jazz). Task 2 asked users to find a song in the collection that was similar to one that was played. The played test songs were sampled from the same collection that was mapped to the browser. The user was asked to repeat tasks 1 and 2 over two phases to better understand how learning of the search space may affect browsing.

In task 1, the tester decided when the correct genre had been found, whereas in task 2 the user decided when they had found a similar song. This was done since it is easy to make broad distinctions between our relatively separate chosen genre classes, using the filenames in our data set as a baseline, and much more difficult to agree upon intra-genre similarity between songs. In task 1, the user was instructed to stop if they had found a song from the correct genre. If a song in the correct genre had not been found, the user was asked to continue searching the space. In task 2 the user decided when a similar song had been found regardless of any baseline metric.

The following questions were asked after the two phases of tasks.

1. How easy was it to navigate the music collection looking for a specific genre?
2. How easy was it to navigate the music collection looking for a similar song?
3. Comment on the difference in your performance between the 2 phases of the tasks
4. Would you want to use this system on a regular basis?
5. Other comments

To better present the outcomes of our informal user study we structured by specific user/ability. Video demonstrations of some of the control interfaces have been also recorded: keyboard ⁴, eye gaze ⁵, iPhone ⁶ and Wii-mote ⁷.

5.1 iPhone and Mouse for a visually impaired user

The first case study was a user with limited vision. He utilized the iPhone as well as the desktop/mouse control interface. In both cases he relied on tactile and audio feedback rather than visual feedback. For the mouse configuration tasks were attempted by navigated to the corners and then used the auditory feedback to guide his search. The search became more efficient as he learned the mapping of genres to the space in the second phase. He found it quite simple to complete the genre task and relatively simple to find similar songs although it did take a longer time. Since he is very familiar with the accessibility features of iTunes on his laptop he would prefer to continue using that. However on a unfamiliar system he thought that the proposed interface would be useful due to its intuitive usage and thought that it would complement iTunes nicely.

⁴http://www.youtube.com/watch?v=DT_mI9Kgtvg

⁵<http://www.youtube.com/watch?v=8eqJ9ZDY5NY>

⁶<http://www.youtube.com/watch?v=ZzrvAt10tiY>

⁷<http://www.youtube.com/watch?v=fhm9QhrYW1U>

Using the iPhone it was easier to orient to the map by using the physical edges of the touch screen. This resulted in faster task completion times on average. He also felt strongly that the proposed system would be more effective for users with limited vision on a personal mobile device with a touch screen.

5.2 Keyboard control on desktop for a motion impaired user with cerebral palsy

In this case the user has cerebral palsy which precludes using his hands for computer usage. He has been trained to use his lips to press keyboard keys. He is fully cognitively and visually abled. In order to navigate the grid that was displayed on a regular desktop monitor the *q,w,a,s* keys were used to go to the top left, top right, bottom left, bottom right corners of the grid respectively. The arrow keys were used for movement and the *x* key was used for centering. Similarly to the previous visually impaired user this participant also initiated each task by going to the corners of the space to orient. Subsequently he used the arrow keys and visual and audio feedback to complete the task. He found the genre tasks very easy to complete, however did have some problems mostly due to lack of knowledge of a particular genre. He easily completed the similarity tasks in one case even finding the exact same song that was being searched for. After learning how the map was organized, the second phase was easier to complete. He commented that he would like this to be integrated into his mobile powered wheel-chair head controls and would also like to have it as part of his personal computer.

5.3 Eye tracker for an able-bodied user

Unfortunately we were not able to recruit a special needs participants for testing the eye tracker and Wii-mote location tracking. To simulate a “locked in” participant for the eye tracking experiment the user was only allowed to move his eyes. Contrary to the other control interface the user did not use the corners for orientation but rather rapidly moved around the space using eye gaze. The user quickly learned the organization of genres upon the space enabling him to efficiently complete the genre and similarity tasks. However during phase II he began to feel eye fatigue since he was not familiar with using eye gaze as a control method. He made the interesting observation that he found it easier to focus his gaze on a particular region when the music was more soft and soothing. Given that he is able to interact in other ways it was hard for him to come to definite conclusions but he did feel that it would be an effective way of interacting for a “locked in” person.

6. CONCLUSIONS AND FUTURE WORK

Browsing of digital music collections is an important activity that is part of daily life for many users. By providing a content-aware representation based on Self-Organizing Maps coupled with personalized control interfaces it is possible to enhance the experience of music browsing and listening for users with special needs such as visual and motor disabilities. We have described the implementation of the systems and initial feedback we received from a small qualitative user study.

There is much future work to be done. We hope to conduct a user study possibly involving multiple concurrent users of the location browsing interface using the Wii-mote

control. More work needs to be done in the initialization of the SOM for personalizing the process to particular music collections. We are also planning on implementing the system on other mobile devices such as BlackBerries and the Google Android platform. Experiments with other clustering and dimensionality reduction methods could be performed. A critical part of the development process will be to incorporate user feedback from users with a range of disabilities. The nature and range of ability of disabled users varies widely, and it will be important to take into account the different needs and abilities of users when planning future developments to this project. Regular usage by users with special needs in their everyday environment will help us get more reliable information for how to improve the system. Some of the suggestions we have received and are working on include: integration with existing music browsing software like iTunes, incorporation of speech recognition and text synthesis for specifying metadata and automatic creation of playlists.

7. ACKNOWLEDGMENTS

We would like to thank CanAssist for providing the motivation and environment for a lot of this work. Telus and the National Science and Research Council of Canada (NSERC) funded parts of this work. Allan Kumka, Jennifer Murdoch and Stephen Hitchner worked on different aspects of the system implementation.

8. REFERENCES

- [1] M. Betke. The Camera Mouse: Visual Tracking of Body features to Provide Computer Access for People With Severe Disabilities. *IEEE Trans. on Neural Systems and Rehabilitation*, 10(1), March 2002.
- [2] R. Bostelman and J. Albus. Sensor Experiments to Facilitate Robot use in Assistive Environments. In *Proc. of the 1st Int. Conference on Pervasive Technologies related to Assistive Environments (PETRAE)*. ACM, 2008.
- [3] C. Doukas and I. Maglogiannis. Enabling human status awareness in assistive environments based on advanced sound and motion data classification. In *Proc. of the 1st Int. Conference on Pervasive Technologies related to Assistive Environments (PETRAE)*. ACM, 2008.
- [4] Y. Ebisawa and S.-i. Satoh. Effectiveness of pupil area detection technique using two light sources and image difference method. *Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE*, pages 1268–1269, 1993.
- [5] O. Goussevskaia, M. Kuhn, and R. Wattenhofer. Exploring Music Collections on Mobile Devices. In *Proc. MobileHCI*, pages 359–362, Amsterdams, the Netherlands, 2008.
- [6] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).
- [7] J. Murdoch and G. Tzanetakis. Interactive content-aware music browsing using the radio drum. In *Proc. IEEE International Conference on Multimedia and Expo(ICME)*, pages 937 – 940, Toronto, Canada, 2006.
- [8] N. Orio. Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, 1(1):1–90, 2006.
- [9] A. Savidis, C. Stephanidis, A. Korte, K. Crispian, and K. Fellbaum. A generic direct-manipulation 3d-auditory environment for hierarchical navigation in non-visual interaction. In *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*, pages 117–123, New York, NY, USA, 1996. ACM.
- [10] G. Tzanetakis. *Marsyas-0.2: A case study in implementing music information retrieval systems*, chapter 2, pages 31–49. *Intelligent Music Information Systems: Tools and Methodologies*. Information Science Reference, 2008. Shen, Shepherd, Cui, Liu (eds).
- [11] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Trans. on Speech and Audio Processing*, 10(5), July 2002.