MusicBox: Navigating the space of your music

Anita Lillie

Thesis Proposal for the Degree of Master of Science in the
Program in Media Arts and Sciences at the
Massachusetts Institute of Technology

December 2007

Advisor:

Tod Machover
Professor of Music and Media
MIT Media Lab

Reader:

John Maeda
E. Rudge and Nancy Allen Professor of Media Arts & Sciences
MIT Media Lab

Reader:

Henry Lieberman
Research Scientist
MIT Media Lab

Reader:

Paul Lamere
Principal Investigator
Sun Research Laboratories

Abstract

Mainstream music browsers have not evolved new methods of organization to deal with growing personal music library size. They instead rely on an exclusively text-based interface, which does not organize music by audio content, nor incorporate the wide variety of user-generated descriptors that is available on the internet. Navigation of music with this type of text-based browser is inherently difficult. This thesis proposal outlines the work required for a new music browser that does incorporate/combine these data, in the form of a three-dimensional map of the user's personal music space. This new interface will be highly-flexible, extendable, and therefore customizable, enabling a fully interactive exploration of that music space, complemented by visually-appealing graphics and informative visualizations. The main goal is that the user will gain a better understanding of his or her musical preferences and listening patterns, while being provided a context for those characterizations. This context also provides the basis for a recommendation system for music.

The Problem

Over the past ten years, widespread use of the MP3 audio encoding format and, in turn, the proliferation of peer-to-peer file sharing and online music stores has lead to a dramatic increase in the average size of personal music libraries [citation]. With increasing library size, navigating, exploring, and finding music has become more difficult, and at times the tasks can seem overwhelming.

Surprisingly, widely available music browsers have not changed their navigation capabilities to keep up with this trend. Commonly-used music browsers *still* rely primarily on lists of static textual metadata (e.g. artist/album/song name, genre, release date, track number), and require the user to recall the connection between the music and these words. They have only limited functionality for personalized navigation based on dynamically-updated metadata, such as play count and track rating. Interfaces like this afford the user little flexibility to organize or search the music in a personal way, nor do they illuminate anything beyond simple relationships between the items in the library.

In addition, music browsers do not provide a way to search the audio content of the music itself, with their “genre” classifier being the only element that allows the user to approximate a search of *what the music sounds like*. Even the “genre” classifier is rife with inconsistency and imprecision; “rock” can describe an enormous range of musical style, for example, and indeed there are many audio tracks that cannot be classified into just one genre.

Here I propose a method which abandons a reliance on these limiting textual classifiers, and instead presents a more abstract, flexible approach to navigating and *exploring* your own music library. This proposal not only provides the groundwork for a new kind of browsing system, but also lays the framework for a larger *recommendation system* in which the user can find *new* music that fits their tastes. Most importantly, mainstream music browsers must adapt to provide this type of innovative and informative interface to keep up with their users’ changing needs.

Proposed Work

I will build a new music browser that organizes a music library’s objects into a 3D space, with object locations determined by multidimensional scaling of similarity measures to other objects in the space. (Here, “objects” refers to whatever level of containment the user wants to apply their music; it could be track, artist, listener, genre, etc.) Similarity measures will be defined by both content-based descriptors (audio) and contextual descriptors gathered from online sources. Traditional metadata associated with each track (including artist name and album art) will be kept and displayed in order to supplement the new data and to ease the transition to the new navigation system.

The term "content-based descriptors" refers here to quantifiable characteristics of the audio content of a piece of music. This could include low-level descriptors like loudness, timbre, pitch, and time segmentation, as well as higher-level descriptors like rhythm, form, and melody. "Contextual descriptors" refers to the listeners' contextual interpretation of an acoustic signal, that is, the words that people use to qualitatively describe musical sound, collected from online reviews, discussion, and tagging of music [1]. Sources for contextual descriptors include these sites: last.fm, allmusic.com, wikipedia.org, youtube.com, and musicbrainz.org.

Similarity models for quantitative (content-based) descriptors will be based upon principal components analysis (PCA), the MARSYAS system [2], and a perceptual music similarity model developed at Sun Microsystems [3]. Models for qualitative (contextual) descriptors are created using co-occurrence analysis.

What results from the music similarity model is a set of data points that will be projected onto a 3D space. Therefore, the user experience of this new music browser is to *enter the space of their music*, and to move about in that space to find the music they seek. Similar objects appear closer together, and the user can see clusters (or segmentation) of albums, artists, songs, and genres. Unlike traditional music browsers, this type of 3D representation of music emphasizes more meaningful *relationships* between different pieces and types of music, and motivates the user to *explore* those relationships. If I am also able to gain access to a larger body of music for analysis, I can create a much larger similarity model, which will allow the user to see the *context* of their musical preferences *inside* the broader perspective of human musical space.

This automatically-generated, three-dimensional model of a user's music library is useful on its own because it creates an intuitively organized space that summarizes a much higher dimensional data set. However, one key focus of this project is to also give the user the ability to *flexibly manipulate and explore* the space, to control and add to what data contributes to the 3D spatialization, and to filter the data based on any component feature. Therefore a bulk of the work in this thesis is directed at building the user interface to enable these sorts of actions.

Here are some of the specific features I plan to implement as part of this user interface:

- Allow user to very flexibly select subsets of songs. These subsets are volumes and paths within the musical space. For example, a user may select an arbitrary-shaped "blob" of tracks, and use it as the base for shuffle playback. A user may also save these volume definitions, for sharing with other users and for personal use in later sessions.
- Playlists can be created by drawing/dragging free-form paths through the space.
- Allow filtering of data based on any component feature. For example, it is easy to visualize only music with a tempo of 120 bpm or higher, or to only visualize music listened to within the past *n* days.

- Allow user to entirely “hide” parameters that are not currently of interest. (e.g. “I don’t care about tempo today. Don’t let it affect the spatialization of my music collection.”)
- Create expandable, “plugin-able” interface: user can import RSS feeds (e.g. from last.fm), or even CSV files, to add more dimensions to the data set.
- Allow user to change the mappings between feature and visual (e.g. timbre and color) to better match their intuitions.
- Live updates of RSS-based data to highlight what user's friends are listening to now in the space.
- Add visualizations of individual song’s inner structure when user zooms in appropriately far. These visualizations will borrow the format used in my previous project, ***soundsieve*** [4].

Because the proposed three-dimensional visualization directly reflects similarity between pieces of music, this browsing tool lays the framework for a music recommendation system as well. As stated previously, if I am able to gain access to a larger body of music, I can show the user their music in the context of music they do not yet listen to. The user can then find music that is similar to music they already like, or, alternatively, find music that fills a “hole” in their collection space. In addition, the static image of one’s listening patterns and preferences within the context of a larger, defined set of music can be interpreted as a *symbol of musical identity.*

I am fortunate to have access to code that can help with this content-based and contextual analysis. The Echo Nest, founded by Media Lab alumni Tristan Jehan and Brian Whitman, has provided me with their *ENCLIAnalyzer* tool, which is available under public licensing agreement [5]. This tool provides quantitative measures of low-level audio descriptors (such as loudness, timbre, and pitch) for each sound segment in any MP3 track [6]. In addition, Paul Lamere at Sun Microsystems Laboratories (Burlington, MA) has agreed to informally share code that: builds content-based music similarity models for a music collection; uses that model to position songs in 3D space with multidimensional scaling [7]; crawls the web for contextual data (i.e. community metadata) from a wealth of sites including last.fm, allmusic.com, and wikipedia.org; and that calculates artist similarity measures based on last.fm tags and last.fm users. Sun Labs intends to release this code as open-source in the future, but there is no set date; therefore any work I build upon this code is only fit for public release contingent on Sun’s release of their own code.

Contribution

This project is novel in that it creates a 3-dimensional model based on audio *and* contextual analysis of a music library, with a user interface that allows flexible manipulation of, and addition to, the visualization of the musical space. All analysis of the musical and textual data is done automatically, requiring no energy on the part of the user to tag or classify any track. At the same time, it illuminates the context for organization and recommendation, with the main goal of enabling the user to *better understand* their music and their musical preferences.

Previous projects/products to analyze, organize, and/or recommend music have generally not attempted to combine both audio and textual data. Projects such as Sun Lab's *Search Inside The Music* have focused only on audio-based data [7], and ultimately concluded that overall *music* similarity is not suitably reflected by *audio* similarity alone [8]. At the same time, purely context-based projects that rely on collaborative filtering have their own concrete limitations. Last.fm and iLike.com make recommendations for new music based on co-occurrence of listeners' track listings and co-occurrence of descriptive tags; amazon.com makes recommendations for new music based on other users' buying histories. But this manner of recommendation requires a (buying or listening) history that already exists, and therefore cannot be applied to brand new music or to music with a tiny (or non-existent) audience.

The European IST project *CUIDADO* contained a "Music Browser", developed by the Sony Computer Science Laboratory, that does incorporate both audio and cultural data, but the interface is riddled with scrolling lists of text descriptors, even when demonstrating artist similarity [9,13]. The *CUIDADO* project has now become the *SemanticHIFI* project, which has outlined some novel interfaces; none of those, however, are similar to the interactive 3D interface put forth here [13]. *SemanticHIFI*'s Music Browser aims at allowing users several means of music access, ranging from completely exploratory to non-exploratory [13]. The project proposed here has those very same goals, yet will apply a *graphical* interface, with searching and navigation accomplished with *movement* (through the space) instead of text-based queries.

Pandora Radio [10] is another interesting example of a confluence of audio and textual data, specifically designed for music recommendation. Pandora's approach to recommending music is to have expert listeners "carefully listen" [10] to and categorize over 400 musical attributes [11] for each song. These attributes are both objective qualities of the music itself (e.g. "Female Vocal", "Prominent Horns", or "Minor Key Tonality"), and subjective characterizations of the music or lyrics (e.g. "Amazing Vocal Prowess", "Joyful Mood", and "Gangsta Rap Attitude") [11]. While this is certainly an interesting approach, it can be so subjective as to be unreliable (it's hard to reach a consensus even between expert listeners), and it is not scaleable to arbitrarily large bodies of music. Because Pandora is likely to only spend time analyzing albums/artists that are relatively popular, they ignore bands with little or no audience base. This approach, like last.fm and iLike.com mentioned above, neglects to address the "long tail" of music [12]. My aim here is to use a *programmatic* approach to audio classification, automatically characterizing musical tracks from an arbitrarily large collection of music without regard to popularity.

This will be the first graphical music browser in which both audio and contextual components are represented, and whose other main focus is to enable such a highly-customizable, information-motivated exploration of one's music collection.

Some questions that can be addressed with an interface that allows such flexible visualization:

- How are different genres related? What kinds of sounds or descriptors characterize these classifications?
- What do the most popular pieces have in common? What do musical pieces that withstand the test of time have in common?
- To what music does a user listen over and over again, and what does that music have in common? What are the audio or textual features that consistently appear in this set of music?
- How varied is the body of music that person *X* listens to? How varied is the body of music that artist *Y* creates?
- How are different remixes, live performances, and productions of the same piece distributed in the audio space?
- How has music evolved over time? How can we better characterize the developments in music over time, specifically those that lead to new "genres"?
- How has an artist's style changed over time?
- If, for example, we have time-based personal listening data (available from last.fm):
    - How have my music listening patterns/habits changed over time? Can I look at the visualization of these trends as a story of what's happening in my life?
    - How has a group or community's listening patterns changed over time?
    - What music do I listen to at different times of day?
    - When did I discover new style *Z*?
    - Why/when do people get "stuck" with their listening preferences at some point in life? Is there a time during life that one is introduced to particular musical patterns that influence one's listening preferences for the rest of his/her life?

Evaluation

In order to test the effectiveness of this new music browser, I will conduct user studies with COUHES review and approval. I will ask users to complete tasks normally done with a music browser (e.g. find a piece that fits your current mood, make a playlist for a particular occasion), as well as tasks that reach beyond the normal scope of a browser (e.g. make a playlist of all your highly rhythmic music, find the five pieces you listen to most in the mornings); then I will compare times and methods for completing these tasks. In addition, I will have users give subjective feedback in the form of a survey.

Schedule

Task	Nov	Dec	Jan	Feb	Mar	Apr	May
Collect and become familiar with resource code	■						
Develop 3D model of music collection		■	■				
Test controller for 3D space		■					
Develop display filters			■	■			
Add support for user extensions			■	■			
Refine user interface; tons of testing				■	■	■	
User studies					■	■	
Write thesis						■	■

Reader Biography

Paul Lamere is Principal Investigator for the 'Search Inside the Music' project at Sun Microsystems Laboratories, where he explores new ways to help people find highly relevant music, even as music collections get very large. Paul is especially interested in hybrid music recommenders and using visualizations to aid music discovery. Currently, Paul is focusing on using signal processing and machine learning to learn to automatically apply social tags to music.

Paul is serving on the program committee for ISMIR 2007, the International Conference on Music Information Retrieval, as well as on the program committee for ACM Recommender Systems 2007. Paul also authors 'Duke Listens!' a blog focusing on music discovery and recommendation.

Paul has an MS in Computer Science from Boston University, and a BA in Chemistry from St. Anselm College.

Bibliography

[1] Brian Whitman. *Learning the meaning of music.* Ph.D. Thesis -- Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2005.

[2] George Tzanetakis and Perry Cook. MARSYAS: A framework for audio analysis. *Organized Sound*, 4(3):169-75, 2000.

[3] Kris West and Paul Lamere. "A Model-Based Approach to Constructing Music Similarity Functions," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 24602, 10 pages, 2007. doi:10.1155/2007/24602

[4] Anita Lillie. *soundsieve*. http://soundsieve.flyingpudding.com/

[5] Licensing agreement and software available at http://analyze.echonest.com/

[6] Tristan Jehan. *Creating music by listening.* Ph.D. Thesis -- Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2005.

[7] Paul Lamere and Douglas Eck. "Using 3D Visualizations to Explore and Discover Music," In *Proceedings of ISMIR 2007 Eighth International Conference on Music Information Retrieval*, September 2007.

[8] Kris West, Stephen Cox, and Paul Lamere. "Incorporating Machine-Learning into Music Similarity Estimation," *Audio and Music Computing for Multimedia*, 2006.

[9] Pachet, F., Aucouturier, J.-J., La Burthe, A., Zils, A. and Beurive, A. The Cuidado Music Browser : an end-to-end Electronic Music Distribution System. *Multimedia Tools and Applications, 2006. Special Issue on the CBMI03 Conference.*

[10] Pandora Radio: "The Music Genome Project." http://www.pandora.com/mgp.shtml

[11] http://en.wikipedia.org/wiki/List_of_Music_Genome_Project_attributes

[12] Chris Anderson. *The Long Tail.* April 2006. http://thelongtail.com/

[13] *SemanticHIFI* Final Public Report. Viewed November 2007. http://shf.ircam.fr/uploads/media/Final_Public_Report.pdf