MOODSWINGS: A COLLABORATIVE GAME FOR MUSIC MOOD LABEL COLLECTION

Youngmoo E. Kim **Erik Schmidt** **Lloyd Emelle**
Electrical & Computer Engineering
Drexel University
{ykim,eschmidt,lte22}@drexel.edu

ABSTRACT

There are many problems in the field of music information retrieval that are not only difficult for machines to solve, but that do not have well-defined answers. In labeling and detecting emotions within music, this lack of specificity makes it difficult to train systems that rely on quantified labels for supervised machine learning. The collection of such "ground truth" data for these subjectively perceived features necessarily requires human subjects. Traditional methods of data collection, such as the hiring of subjects, can be flawed, since labeling tasks are time-consuming, tedious, and expensive. Recently, there have been many initiatives to use customized online games to harness so-called "Human Computation" for the collection of label data, and several such games have been proposed to collect labels spanning an excerpt of music. We present a new game, MoodSwings (http://schubert.ece.drexel.edu/moodswings), which differs in that it records dynamic (per-second) labels of players' mood ratings of music, in keeping with the unique time-varying quality of musical mood. As in prior collaborative game approaches, players are partnered to verify each others' results, and the game is designed to maximize consensus-building between users. We present preliminary results from an initial set of game play data.

1 INTRODUCTION

The detection and labeling of the emotional content (mood) of music is one of many music information retrieval problems without a clear "ground truth" answer. The lack of easily obtained ground truth for these kinds of problems further complicates the development of automated solutions, since classification methods often employ a supervised learning approach relying on such ground truth labels. The collection of this data on subjectively perceived features, such as musical mood, necessarily requires human subjects. But traditional methods of data collection, such as the hiring of subjects, have their share of difficulties since labeling tasks can be time-consuming, tedious, error-prone and expensive.

Recently, a significant amount of attention has been placed on the use of collaborative online games to collect such ground truth labels for difficult problems, harnessing so-called "Human Computation". For example, von Ahn et al. have created several such games for image labeling: the *ESP Game*, *Peekaboom* [1], and *Phetch*. More recently, several such games have been been proposed for the collection of music data, such as *MajorMiner* [2], *Listen Game* [3], and *TagATune* [4]. These implementations have primarily focused on the collection of descriptive labels for a relatively short audio clip.

We present a new game, *MoodSwings*, designed to explore the unique time-varying nature of musical mood. Of course, one of the joys of music is that the mood of a piece may change over time, gradually or suddenly. According to Huron [5], this combination of anticipation and surprise may be at the core of our enjoyment of music. Thus, our game is targeted at collecting dynamic (per-second) labels of users' mood ratings, which are collected in real-time as a player hears the music using the two-dimensional grid of emotional components: valence and arousal. As in other collaborative games, players are partnered in order to verify each others' results, providing a strong incentive for producing high-quality labels that others can agree upon. Accordingly, game scoring is designed to maximize consensus-building between partners. In this paper, we present data from an initial pilot phase of the game and demonstrate the utility of this approach for the collection of high-quality, dynamic labels of musical affect.

2 BACKGROUND

Models of affect and the categorization and labeling of specific emotions has received significant attention from a variety of research areas including psychology, physiology, neuroscience, as well as musicology. With the advent of digital music and very large music collections, recent work has focused on the problem of automatic music mood detection. Next, we briefly summarize some of the related work.

2.1 Mood models

Early work on the quantification of musical affect focused on the formation of ontologies using clusters of common emotional adjectives and labels (e.g., "bright", "gloomy", "contemplative", "angry"). Ontologies proposed by Hevner [6] and Farnsworth [7] proposed eight and ten such mood clusters, respectively. All Music Guide [8], a large edited music information database, also uses a descriptive approach with a total of 179 distinct (but not unrelated) mood labels.

An alternative approach to the modeling of human emotions views affect as a combination of orthogonal continuous sub-features. The most popular such representation of musical affect is Thayer's two-dimensional valence-arousal space [9], which itself is derived from Russell's general model of human emotions (pleasant-unpleasant vs. arousal-sleep) [10]. Thayer's model decomposes emotion in music according to two principal dimensions:

- *valence:* positive vs. negative (e.g., happy vs. sad)
- *arousal:* high- vs. low-energy (e.g., energetic vs. calm)

According to this model, music can be broadly categorized into one of four quadrants: high valence and arousal (joy, exuberance), high valence and low arousal (contentment), low valence and high arousal (anger), and low valence and arousal (depression). But the model also views the axes as continuous features, allowing for an unlimited combination of overall moods. The continuous valence-arousal model is at the core of MoodSwings.

2.2 Ground truth label collection

The tedium and expense of label collection has presented a significant obstacle for researchers seeking such labels in order to train automatic systems for mood classification. When employing subjects specifically for the collection of mood labels for music, prior systems have used alternatively expert and non-expert populations.

The Pandora service [11] is an example of a well-known expert labeling system that employs a large team of musicians to manually label tracks according to hundreds of features (their "music genome"), some of which relate to mood. The former MoodLogic service used questionnaires given to their users to collect mood metadata. The data from these services is, sadly, not available to the public. Other commercial tools allow users to tag their own collections, such as the Moody plug-in for iTunes, which uses a quantized 4x4 valence-arousal grid.

More recently, online services that allow users to input free-form (unconstrained) tags, such as Last.fm [12], have collected myriad tags (some of which represent mood labels) across very large music collections. The accessibility of this data has led to a recent trend towards using such free-form tags as the basis for mood labeling by aggregating results from many users. Hu, Bay, and Downie collected labels and tags from AMG, Last.fm, and epinions.com to form the ground-truth mood "clusters" used in the 2007 MIREX mood detection evaluation [13].

2.3 Automatic mood classification of music

The general approach to automatic mood detection from audio has been to use supervised machine learning to train statistical models of acoustic features. Li and Ogihara [14] used acoustic features related to timbre, rhythm, and pitch to train Support Vector Machines (SVMs) to classify music into 13 mood categories derived from Farnsworth's emotion groupings. Using a hand-labeled library of 499 music clips (30-seconds each), they achieved an accuracy of $\sim$45%, with 50% of the database used for training and testing, respectively.

Lu, Liu, and Zhang [15] pursued mood detection and tracking (following dynamic mood changes during a song) using a variety of acoustic features related to intensity, timbre, and rhythm. Their classifier used Gaussian Mixture Models (GMMs) for Thayer's four principal mood quadrants in the valence-arousal representation. The system was trained using a set of 800 classical music clips (from a data set of 250 pieces), each 20 seconds in duration, hand labeled to one of the 4 quadrants. Their system achieved an accuracy of $\sim$85% when trained on 75% of the clips and tested on the remaining 25%.

In 2007, the Music Information Research Evaluation eXchange (MIREX) first included a "beta" task on audio music mood classification with 8 systems submitted. The audio clips used for this task were assigned to one of 5 mood clusters, aggregated from AMG mood labels (adjectives), and 600 30-second hand-labeled clips distributed equally among the 5 mood clusters were used in the evaluations. All participants performed reasonably well (far higher than chance) with the highest performing system [16] achieving correct classifications slightly over 60% of the time. It should be noted that several of the systems were primarily designed for the genre classification and then appropriated to the mood classification task as well [17].

3 MOODSWINGS

MoodSwings is a collaborative, two-player game that incorporates each listener's subjective judgements of the emotional content (mood) of music into the game play. At the start of a match, a player is partnered with another player anonymously across the internet. The goal of the game is for the players to dynamically and continuously reach agreement on the mood of 5 short (30-second) music clips drawn from a database of popular music.

The MoodSwings game board (Figure 1) is a direct representation of the valence-arousal space. The board represents a continuum of possible mood ratings, with arousal on the horizontal axis and valence on the vertical axis. During gameplay, players simultaneously listen to identical short music clips. Each player positions their circular cursor dynamically on the game board, indicating their instantaneous

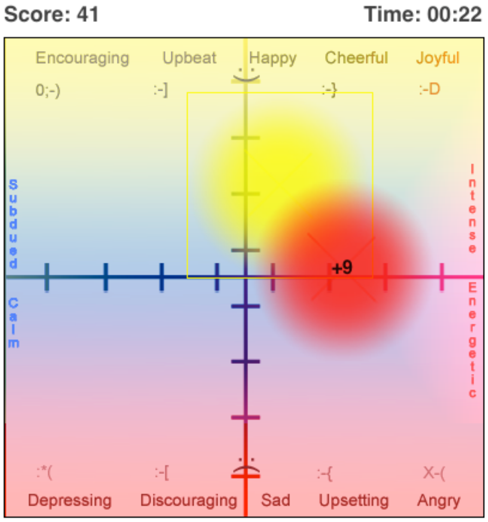


Figure 1. The MoodSwings game board

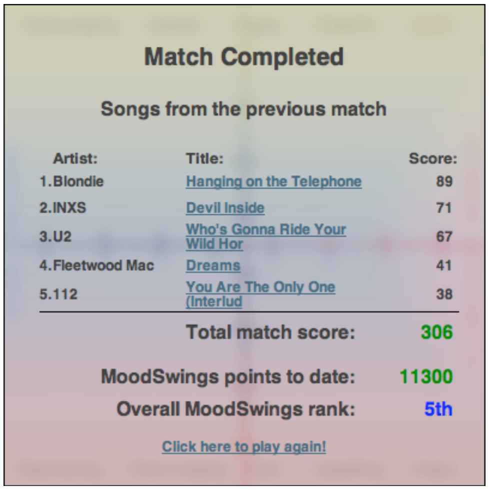
assessment of the mood of the music. A player's position is sampled once per second, indicated by the pulsing of their cursor. The partner's cursor is visible only intermittently, every few seconds. Scoring is based on the amount of overlap between the partners' circles, with greater congruency resulting in a greater number of points scored. The size of the players' cursors decreases over time as the clip plays, increasing the difficulty of scoring points as time elapses (the players must agree on their ratings more precisely to overlap and thus score points).

3.1 Game play sequence

A MoodSwings match consists of 5 rounds, each consisting of a different music clip with a duration of 30 seconds. Once a partner pairing is established and verified, each round commences after a short 3-second countdown. The partner pairing remains consistent for all 5 rounds. The game board remains inactive until the round begins.

1. Once the round begins, the player's cursor (colored yellow circle) becomes visible. The cursor "throbs" every second indicating the sampled position, but the player is free to continuously alter the cursor position between pulses.
2. After an initial period of 5 seconds, the partner's cursor becomes visible for the first time (again pulsing for one second). This allows a player to make an initial mood assessment independently, without influence from their partner.
3. Afterwards, the partner's cursor is visible once every 3 seconds. This interval is designed to prevent players from simply "chasing" their partners in order to accumulate more points.
4. The size of both cursors decreases continuously during the course of the round.

At the end of each round, the player is presented with their score for the round and their total points for the match thus far. Likewise, at the conclusion of a match, a list is presented with performing artist and title for each music clip (offering a link to a search engine for each piece), as well as the points for each round and the total score for the match (Figure 2).



Match Completed

Songs from the previous match

Artist:	Title:	Score:
1.Blondie	Hanging on the Telephone	89
2.INXS	Devil Inside	71
3.U2	Who's Gonna Ride Your Wild Hor	67
4.Fleetwood Mac	Dreams	41
5.112	You Are The Only One (Interlud	38

Total match score: 306

MoodSwings points to date: 11300

Overall MoodSwings rank: 5th

Click here to play again!

Figure 2. The MoodSwings post-match wrapup

3.2 Game scoring

The primary score calculation is based upon the amount of overlap between the player's cursor and that of their partner, which is designed to encourage maximum agreement in the mood assessments from both parties. This score is only accumulated whenever the partner's cursor is visible. When points accumulate, they are displayed on top of the partner's cursor (shown in Figure 1).

Players can also accumulate *bonus points* by "convincing" their partner to agree with a particular mood assessment (position). This provides an incentive for players to remain firm in a position and not to be capricious in their movements. It also encourages them to respond to a change in mood rapidly in order to "stake out" the new position before their partner. The rules for bonus points are as follows:

- A player is eligible to accumulate bonus points after remaining stationary for 1 second. This is indicated by a yellow square around the player's cursor. Bonus points will only be awarded while the player remains stationary.
- If a partner moves towards a player's location achieving overlap between the cursors, the stationary player

is awarded 5 bonus points. This is indicated by the cursor changing color to green (Figure 3).

- Bonus points may be awarded every second, even when the partner's cursor is not visible to the player.

The overall point system is designed so that a "good" score for a round is approximately 100 points, and a particularly good match total is 500 points. High scores for the top five players as well as the top five individual match scores are visible when first logging onto the MoodSwings website.

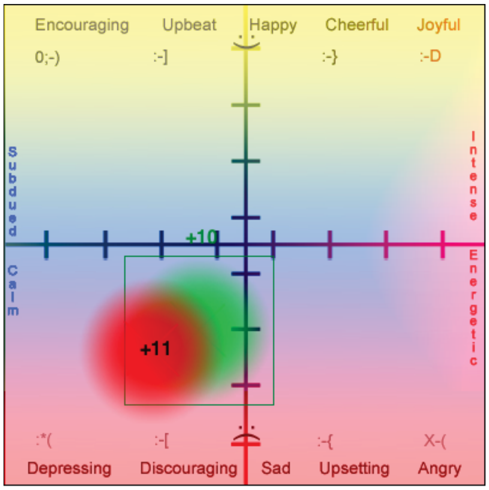


Figure 3. MoodSwings bonus scoring

3.3 Music database

The music for MoodSwings is drawn randomly from the well-known uspop2002 database, a collection of over 8000 popular music tracks from approximately 400 performing artists [18]. This database was chosen because of the composition of the corpus (popular music spanning several decades) and its potential appeal to a mass audience, the size of the database, and the fact that it includes a range of familiar and obscure tunes. In addition, since the corpus is well-known, a great deal of acoustic feature data and music metadata have already been calculated and compiled for the database, such as MFCCs [19]. This will allow other researchers to rapidly and easily deploy their algorithms using the labels collected from the game.

The uspop2002 database, however, is not without its issues. A fair number of the songs contain explicit lyrics, which may be objectionable to some players. Some of the clips randomly selected for the game will not include music at all because of extended applause from live recordings and other unusual spoken-word tracks from a few albums. Because such metadata (explicit lyrics, non-music sections) for tracks is not readily available, the game interface also offers players the opportunity to mark such tracks. In particular, beneath the game board for each round are the following options that a player may voluntarily select:

- *Clip does not contain music:* for selections that do not contain music (so that we can filter these tracks out in the future).
- *Song contains explicit lyrics:* this will help us create a future version of the game appropriate for all ages that excludes these songs.
- *Report a bug in this game:* for any other problem encountered during the current round.

3.4 Technical implementation

A primary goal during the development of MoodSwings was to allow access to the game through a standard web browser interface, so that the game would be widely accessible and would not require the installation of any additional software. In particular, our lab works closely with several K-12 school environments where the computers may be out of date and additional software is difficult to install.

The end-user interface for MoodSwings is principally coded using Asynchronous JavaScript and XML (AJAX), which also makes use of Dynamic HTML. Audio playback is handled through the Flash plug-in, which is included with all browsers. The browser portion communicates with a web server hosted by our laboratory that is used to synchronize players, record the collected mood assessments, and administer the game. A diagram of the overall architecture of the system is given in Figure 4.

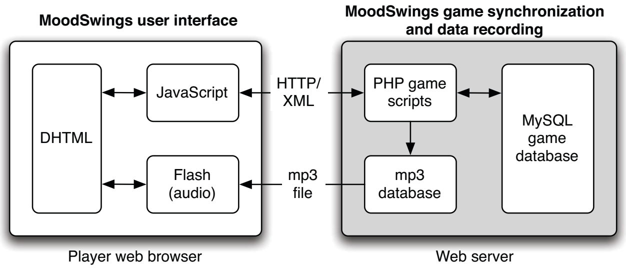


Figure 4. MoodSwings game architecture for a single player. Each player interacts with the server independently.

The web server uses scripts written in PHP to respond to client requests for game data, such as music clip information and partner mood coordinates, in XML format. The game data is stored in a MySQL database running on the server that tracks all of the following information:

- Valence and arousal assessments from each player for each second of the song.
- Game and match scores.

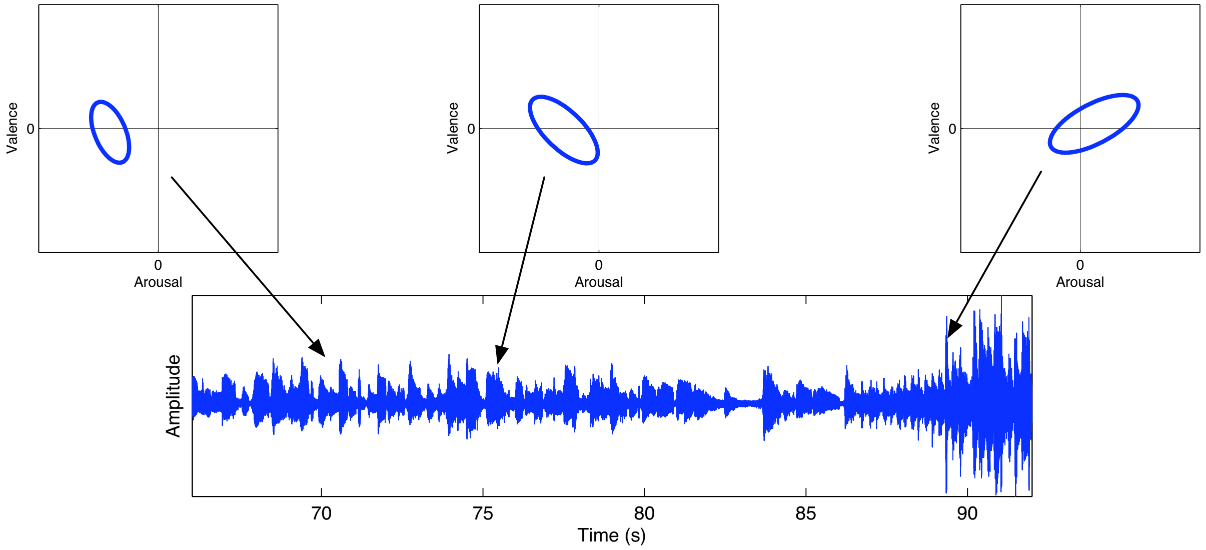


Figure 5. Progression of valence-arousal labels over a clip spanning the end of the first chorus and beginning of the second verse of "American Pie" (Don McLean). The ellipses represent the standard deviation across different players.

- Game and match parameters (song clips used, clip starting locations).
- Music metadata (artist, album, song title).
- User scores and high scores.

3.4.1 Technical limitations and single-player matches

Obviously, with simultaneous listening between the players there will be some latency in the data received from one's partner during a game. Since the mood assessments are only sampled once per second, it is likely that the partner's data received by a player will be at least one second out of date. In practice, this does not affect gameplay very much, since a player normally takes a second or two to react to sudden mood changes.

Even when other players are not available online, a player may still play a match against data previously recorded from another player's match. No distinction is made to the player to indicate a match against a live player vs. a pre-recorded match. In some ways, such single-player matches against recorded data may be preferred because the partner's labels can be pre-fetched so that synchronization between ratings and time samples is not an issue.

4 MOOD DATA COLLECTION

In our initial one week pilot phase, we attracted approximately 100 users and collected over 50,000 valence-arousal point labels spanning more than 1000 songs. Of course, given the collaborative structure of the game, many of the valence-arousal labels refer to the same locations in a song.

4.1 Example data from song clip in database

Figure 5 shows a sample of collected valence-arousal labels for a short section of the song "American Pie" by Don McLean, between the first chorus and second verse. The change in instrumentation and tempo within the segment (drums are added to voice and guitar) is generally marked by players as a change in intensity in the song, as well as a slight increase in valence.

4.2 Summary of labels collected to date

Figure 6 depicts the distribution of the collected points in the valence-arousal space. It is clear that the songs labeled thus far have bias towards high valence and arousal, which is consistent with a database of popular music. The plot also shows that players for the most part favor locations near the middle of each "quadrant" in the valence-arousal space, largely avoiding extreme values.

5 FUTURE WORK

The initial implementation of MoodSwings and the preliminary data collected thus far suggest the potential of such a collaborative system with many users in providing high-quality labels. We are currently investigating modifications to make the gameplay more fun, with the hope of drawing a greater number of repeat visits from users. A question arises in examining live matches vs. those played against recorded data. Qualitatively, the labels collected in single and two-player games appear equally valid, but we plan to verify their consistency by more closely examining the data.

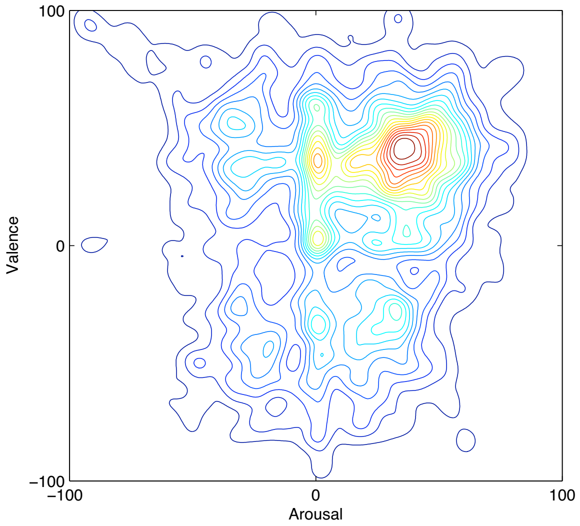


Figure 6. Contour plot of the distribution of ∼50,000 valence-arousal labels collected over one-week pilot period.

An issue in dealing with popular music is that the mood of most songs simply doesn't change very much or often, which can lead to rather static games (perhaps this is a comment on the state of popular music!). We plan to add an option for additional simultaneous players (3 or even more) that may produce more interesting group labeling dynamics. We also plan on augmenting the uspop2002 databse with a collection of classical music clips, which will have the added benefit of removing some of the lyric vs. audio confusion that arises in labeling some pop songs.

As our collection of labels grows, we intend to use the data to train a system for mood classification of short audio segments using a Hidden Markov Model to capture time-varying mood transitions. Once the entire database has been labeled, the data collected from MoodSwings will be made available to the music information retrieval research community through the game website.

6 REFERENCES

[1] L. von Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, 2006.

[2] M. I. Mandel and D. P. W. Ellis, "A web-based game for collecting music metadata," in *Proceedings of the 8th International Conference on Music Information Retreival*, Vienna, Austria, 2007, pp. 365–366.

[3] D. Turnbull, R. Liu, L. Barrington, and L. G., "A game-based approach for collecting semantic annotations of music," in *Proc. 8th International Conference on Music Information Retreival*, Vienna, Austria, 2007, pp. 535–538.

[4] E. L. M. Law, L. von Ahn, R. B. Dannenberg, and M. Crawford, "TagATune: a game for music and sound annotation," in *Proc. of the 8th International Conference on Music Information Retreival*, Vienna, Austria, 2007.

[5] D. Huron, *Sweet Anticipation: music and the psychology of expectation*. Cambridge, MA: MIT Press, 2006.

[6] K. Hevner, "Experimental studies of the elements of expression in music," *American Journal of Psychology*, no. 48, pp. 246–268, 1936.

[7] P. R. Farnsworth, "The social psychology of music," *The Dryden Press*, 1958.

[8] "The All Music Guide." [Online]. Available: http://www.allmusic.com

[9] R. E. Thayer, *The Biopsychology of Mood and Arousal*. Oxford, U.K.: Oxford Univ. Press, 1989.

[10] J. A. Russell, "A complex model of affect," *J. Personality Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[11] "Pandora." [Online]. Available: www.pandora.com

[12] "Last.fm." [Online]. Available: http://www.last.fm

[13] X. Hu, M. Bay, and J. S. Downie, "Creating a simplified music mood creating a simplified music mood classification ground-truth set," in *Proc. 8th International Conference on Music Information Retreival*, Vienna, Austria, 2007, pp. 309–310.

[14] T. Li and O. M., "Detecting emotion in music," in *Proc. 4th International Conference on Music Information Retreival*, Baltimore, Maryland, 2003.

[15] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 5–18, 2006.

[16] G. Tzanetakis, "Marsyas submissions to MIREX 2007," MIREX 2007.

[17] J. S. Downie, "The 2007 MIREX results overview." [Online]. Available: http://www.music-ir.org/mirex2007/

[18] A. Berenzweig, B. Logan, D. Ellis, and B. Whitman, "A large-scale evaluation of acoustic and subjective music similarity measures," *Computer Music Journal*, vol. 28, no. 2, pp. 63–76, June 2004.

[19] D. P. W. Ellis, A. Berenzweig, and B. Whitman, "The *uspop2002* pop music data set." [Online]. Available: http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html