

SmartDJ, An Interactive Music Player for Music Discovery by Similarity Comparison

First author

Aw Si Ying Maureen
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
maureenaw@outlook.com

Second author

Lim Chung Sion
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
limcs@me.com

Third author

PerMagnus Lindborg
Nanyang Technological University
50 Nanyang Avenue
Singapore 639798
permagnus@ntu.edu.sg

ABSTRACT

In this digital music era, sorting and discovery of songs is getting harder and more time consuming than before, due to the large pool of songs out there. Many music recommendation system and other similar applications in the market make use of collaborative filtering and social recommendation to suggest music to listeners. However, the problem arises when there is not enough information collected for the song, which happens mostly to new and less popular music. Other issues include missing or inaccurate metadata, the need for Internet connection, etc.

We present research on acoustic features to automatically classify songs according to user-friendly and high-level concepts that indicate social contexts for music listening, and a prototype application called "SmartDJ". We aim to provide novel ways that the user can browse her/his music collection, with a player that enhances interaction via a visual feedback, personalised DJ trajectories, smooth mix transitions and so forth. SmartDJ sorts the songs based on similarity by extracting low level features, then reducing feature space dimensionality with principle component analysis (PCA) and multidimensional scaling (MDS) methods, and plotting songs in a GUI for manual or automatic browsing, where song similarity is given by Euclidian distance in a lower-dimension song space. Users are able to visualise their music library and select songs based on their similarity, or allow the system to perform automation, by selecting a list of songs based on the selection of the seed song. Users can maneuver with the high-level descriptor on the interactive interface to attain the different song space desired.

1. INTRODUCTION

Music discovery system is essential for users to explore songs from a large collection of music. The idea of SmartDJ is to serve as a personal Deejay (DJ) to make the selection of song choices for users without having the skillset of a DJ. The system automatically generates a playlist of songs based on the seed song and/ or user can make selection of songs, all based on song similarity. We proposed a new and interactive way of visualizing a personal music library by translating all the songs into a song space to provide a form of visual feedback. The similarity between the songs is determined by its proximity.

In order to achieve the song similarity comparison, signal analysis is performed on individual song. Low-level descriptors are extracted for similarity measurement. The

large dataset is then reduced with the use of dimension reduction techniques such as principle component analysis (PCA) and multidimensional scaling (MDS) methods, for easy viewing by users.

The song space model can be adjusted accordingly with different inputs from the user to suit the different scenarios or needs.

2. BACKGROUND

A well-know model that is applicable to the case of our song space model is Thayer's mood model (1989) as depicted in Figure 1. Thayer's mood model divides mood into two allegedly uncorrelated dimension vectors: arousal and valence [1]. Arousal can be described as the energy or activation of an emotion. Low arousal corresponds to feeling sleepy or sluggish while high arousal corresponds to feeling frantic or excited. Valence describes how positive or negative an emotion is. Low valence corresponds to feeling negative, sad or melancholic and high valence to feeling positive, happy or joyful.

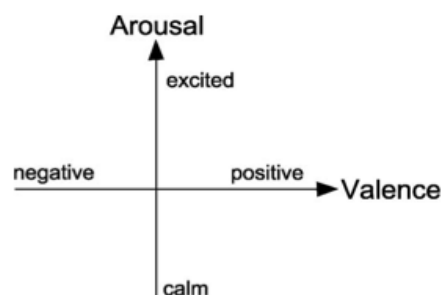


Figure 1. Thayer's Mood Model

Microsoft Research Asia (Liu, Lu, et al., 2003) proposed a method to properly use Thayer's model for music mood classification, in which mood was divided into four nominal classes resembling the four quadrants in the mood plane spanned by the two vectors. The first quadrant (excited & positive) corresponds to 'happy/ excited' emotion, second quadrant (excited & negative) corresponds to 'angry/ anxious' emotion, third quadrant (calm & negative) corresponds to 'sad/ bored' emotion and finally the last quadrant (calm & positive) corresponds to 'relax/ serene' emotion. This model is further elaborated with the various emotions as labeled on an Arousal-Valence (A-V) space [2] shown in Figure 2. We can aim to apply Thayer's mood model to our song space model. As such we will have a better feel of how the different songs are position in the space. Hence, manipulating the

song space with different user input and the system can then select songs from the right space to suit the needs of the user.

Aside from mood model, many also suggest other forms of song classification, genre classification in particular. Davalos [3] suggests using Linear discriminant analysis for dimension reduction so as to project the data for optimum class separation. While Clark, Park and Guerard [4] suggested Growing neural gas (GNG) as a form of self-organizing map, Langlois and Marques [5] suggested Hidden Markov Models (HMMs) for genre classification. Keeping in mind that our objective is to project songs into a song space based on similarity and not based on genre or other features, the methods suggested can be explored and used as a form of reference for our work.

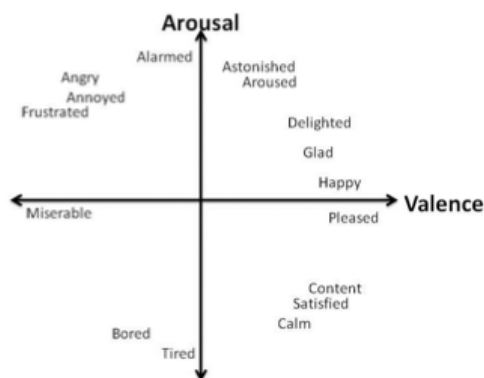


Figure 2. A-V Space Labeled with Different Emotions

3. PLATFORM

Matlab R2010b [6] serves as the main platform for developing the prototype of the song space model. Dimension reduction techniques such as PCA and plotting of the two-dimensional song space were performed with the use of Matlab. MIRtoolbox 1.4 [7] is an essential tool that rides on Matlab and is used for low-level features extraction in our work.

Max 6 (MSP) [8] is used to develop the GUI for our work, which involves interface design, data organization and processing. The working model developed from Matlab will be ported over to Max/MSP to serve as the backbone for our application, and will work hand-in-hand with the user interface. A third party component, ircam-descriptor [9] is a real-time descriptor analyser for Max/MSP. It is capable of performing features extraction and other signal processing analysis offline. Therefore, this allows songs to be imported to SmartDJ for analysis and to plot them onto the song space, without having the need of playing the songs unlike in the case of a real-time analysis.

4. SONG SPACE DEVELOPMENT

Similarity between songs is a subjective measure. Many software deals with this by defining based on a certain genres, artists, etc. And more often recommend songs to user based on the popularity of the song (play count) and

by collaborative filtering, which means that, if listeners who like song A, B and C also like song E, then the system is likely to recommend song E to other listeners who listen to song A, B and C.

In the case of SmartDJ, it sorts out the similarity of songs by first extracting low-level features, such as brightness, centroid, roll-off, Mel-Frequency Cepstral Coefficient (MFCC), etc. This in total makes up 28 features, which is inclusive of the 13 MFCC coefficients. All features were extracted with the use of MIRtoolbox 1.4.

4.1 Features Extraction

A corpus of 310 songs was used in the training dataset. The audio files are in lossless WAVE format, encoded in linear pulse code modulation (PCM) of 16 bits in stereo channels with a resulting audio bit rate of 1411200 bit/s. Only an excerpt of 30 seconds of the middle segment of the songs was examined. The middle segment of the song was a sensible choice, as intuitively, it is where the gist of the song is, and in most cases it is the chorus of the music. Even though this may not always be the case, but most of the time true. Davalos however chose to analyze the first 30 seconds [3] of the song. The audio signal was then down-sampled to 22050Hz [10], which is similar to the case of Arenas-Garcia, Petersen, and Hansen (2007). Even though the experiment was conducted in an ideal scenario, but in actual fact during implementation, users might be more prone to mp3 files due to its smaller file size and compactness. But further investigation will have to be done to determine if the end result will be affected, which will not cover in this paper.

4.2 Dimension Reduction

The large dataset collected from the corpus is then reduced in dimension and projected onto a two-dimensional song space as a form of visual feedback to the user. The similarity between the songs can be determined from the plot with similar songs being situated near each other and songs that are very different being plotted far away from one another. In order to reduce the dataset into a two-dimensional plot, dimension reduction techniques have to be employed. In our case, we chose principle component analysis (PCA) and multidimensional scaling (MDS) methods, where song similarity is given by Euclidian distance in a lower-dimension song space. With PCA, the highest variance is retained in the first and second principle component (PC) respectively, giving it the greatest spread. Hence, as shown in Figure 3, the data is plotted with PC2 against PC1, in order to retain most of the information.

With the 28 features, the percentage variability for the first two and three PCs account for only 24.99% and 32.43% respectively, which is fairly poor. To overcome overfitting issue and to improve the variability explained, the features employed were further streamline to nine spectral shapes features. The song space model obtained from spectral shape features is shown in Figure 3.

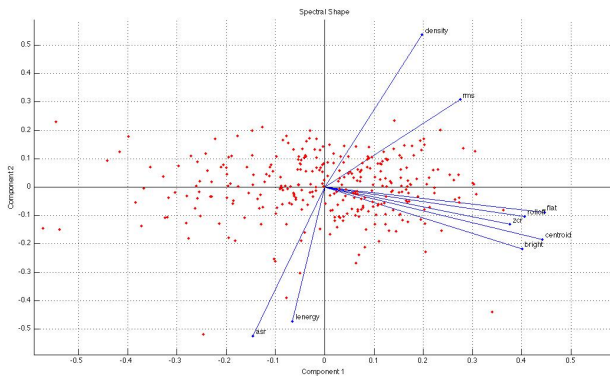


Figure 3. Song Space Model with spectral shape features.

Results from the first three PCs account for 51.05%, 72.87% and 81.64% (See Figure 4) of the variance respectively. A sharp bend at the second PC indicates that the variability explained by the third PC onwards is not as significant. Hence, a two-dimensional plot with the first two PCs is employed in our model.

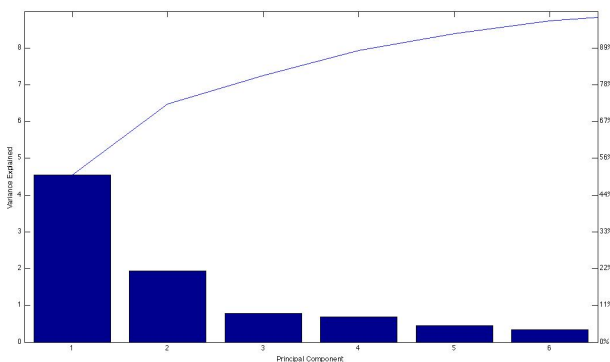


Figure 4. Percent variability explained by the first 6 PCs.

4.3 Song Space Model

A general trend of the song placement can be noticed from our song space model. The typical Pop, Rock and Techno songs take up the first quadrant (upper right). We would consider this quadrant to contain songs with higher danceability and are generally high, bright and nosier. Songs that are dark but rich in audio content takes up the second quadrant (upper left) and they are songs that are generally more Acoustic and Country. Songs from artists such as Taylor Swift, Jason Mraz and Bruno Mars tend to appear more often in this quadrant. Music that is dark and generally more melancholy takes up the third quadrant (lower left). They are songs that are more instrumental and jazzier for example songs from artists like Adele and Kenny G tend to appear in this quadrant. Finally, the fourth quadrant (lower right) contains songs that are generally melancholy but with a faster beat. Examples include Jazz and Country music with a hype. This division into the four quadrants draws back the relation to Thayer's mood model (Figure 1), which similarly categorizes the mood plane into four main sectors spanned by the two vectors. This suggests that the song space can be classified in terms of emotional mood and user can select

the type of music on the song space based on their social context.

From the general trend observed above, we observed that as song progresses along the horizontal axis, it moves from a “quieter” zone to a “nosier” zone. And in terms of genre, this means that songs change from Jazz, Acoustics and light-hearted Country songs to Rock, Techno, Pop and House music. Thus, x-axis (PC1) corresponds to the noisiness of a song. The vertical axis increases in energy level as it progresses from bottom to top. Jazz music is generally located at the bottom, while Pop music is generally located at the upper half of the song space. Thus, y-axis (PC2) relates to the massiveness or heaviness of a song. This analysis is supported by the features' loadings as shown in Figure 3. It can be seen that the loadings for features like flatness, rolloff, zero crossing rate (ZCR), centroid and brightness lie closer to the horizontal axis, and they measure the amount of high frequency energy and how much the signal oscillate. Thus it corresponds with our analysis by saying that the horizontal axis is a measure of how noisy, or saturated with high-frequency content, the music is. Also the loadings of density and RMS point in the direction close to the positive y-axis while loadings of low energy and absolute silent ratio (ASR) point in the negative y-axis direction. Hence, corresponds with our analysis with the vertical axis being a measure of the amount of energy or how *massive/ heavy* the music is.

5. SMARTDJ INTERFACE

The idea of interface design for SmartDJ is focused on non-DJ users who utilise SmartDJ as their daily music player without too much hassle. The simplicity and interactivity is the design focus for SmartDJ.

5.1 Soundbar Mode

Soundbar Mode allows the user to simply play tracks with minimum action required. The Soundbar mode comes with three main panels for basic operation. SmartDJ panel provides trigger buttons for effect automation such as auto-crossfading and party mode. Visualiser, playlist editor and advanced setting are located under Features tab for triggering the pop-up window according to user's need.



Figure 5. Soundbar Mode

The main music player control includes essential controls and information to provide the ease of use to amateurs. Essential controls such as previous song, play, next song, loop, shuffle and additional information provide basic control for the music player. Information such as current playing track information and upcoming track are included to provide information feedback to the users. Upcoming track information tab that paired with next button to provide instant reselection for user who is not satisfy

with the next song. He/she can change the upcoming track until he/she finds the desired song to cue.

5.2 DJ Mode

DJ mode is created to simulate the interface of conventional DJ software with simplified features for the non-DJ users such as cross-fader, volume, speed, pitch and parametric equaliser for individual play desk to perform manual manipulation. Users are able to achieve beat synchronization manually but changing the speed of both deck individual to match the song speed of both. Furthermore, the pitch of song can be maintained as original by manipulating the pitch slider. These manual features give better interaction without increase the difficulty of using SmartDJ.

Effects such as beat synchronisation and song structure detection can be applied on DJ mode. Beat synchronisation helps to match the speed of song for both tracks for the ease of song transition where it gives better song transition effect when both BPMs of song are nicely matched. Key lock feature allows the system to maintain the pitch of song while changing the speed. Song structure detection is the concept applies for finding best mixing points for users. Song structure detection separates song into different part such as verse, chorus, intro and outro. This information can be used for users to perform manual crossfading at desired point or serves as reference for the automation to pick the appropriate point for mixing. Currently, this component is considered as a part for future development.



Figure 7. DJ Mode

6. INTERACTIVE FEATURES

The interface design for SmartDJ is focused on non-DJ users who utilise SmartDJ. The objective of developing SmartDJ is to create a new way of interaction between the users with music player. Therefore, interactive features are the main focus of the SmartDJ development. Song Space Visualiser and Smart Equaliser are the main features that emphasise on user interactivity. Advance Setting Panel provides further adjustment for the system to suit the users' need better.

6.1 Song Space Visualiser

Song Space Visualiser creates a new way of interaction between user and music player by providing a visual feedback regarding to the song similarity of songs that added into SmartDJ.

The analysed result after dimension reduction will then plot into smaller dimensional song space. This visual feedback helps the users to understand that what are the

similar songs around the seed song. Alternatively, SmartDJ is able to select the subsequence songs based on seed song automatically if the users choose to activate the automation.

This is inspired by the work of CataRT [11] and MusicBox by Anita Lillie [12]. CataRT is a real-time sound synthesis system that allows display the corpus of songs data on a space based on its proximity in descriptor space. The concept of MusicBox is one step further from CataRT and closer to our idea. It projects a large corpus of songs in a space and the model can be adjusted by filtering different descriptors. Song Space Visualiser aims to improve the idea into a potential application that comes with higher usability, higher user interactivity and higher user friendliness.

The 3 different interfaces then evaluated by group of people who are amateur users that use music player very often. They verify these programmes based on the accuracy of the presentation, user interactivity and user friendliness. The result is SmartDJ scored the better overall score in terms of user interactivity and user friendliness.

In SmartDJ, there are 2 song spaces, which displaying the relationship of BPM against key of songs (Song Space 1) and the relationship of song similarity (Song Space 2). These 2 song spaces are formed by 5 dimension data, which include BPM, key, Brightness, Noisiness and Heaviness. With different dimension arrangement, these 2 song spaces provide different visual feedback to the user.

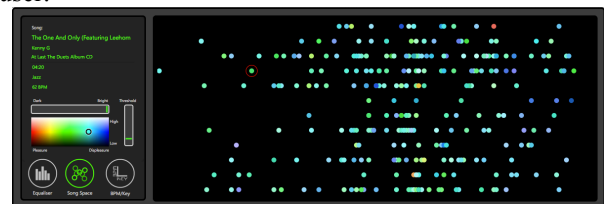


Figure 1. Song Space Based on BPM and Key

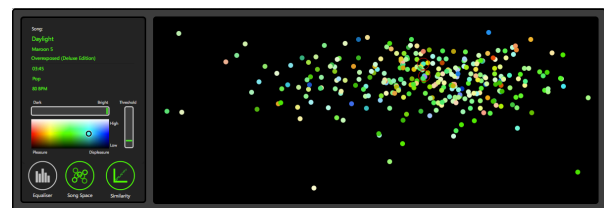


Figure 2. Song Space Based on Similarity

In the background, K-th Nearest Neighbour used to choose songs that are closest to the seed song with the inputs provided by the user. User can define the number of song selection as well to include more songs for the system to perform filtering. The system will have a higher chance to locate the best candidate to prevent the system falls into a loop where the seed song and selected song will keep playing repeatedly.

In order to overcome this issue, SmartDJ introduces *Social Input* as input parameter of song space for user to define direction for the song selection system to select subsequent songs based on seed song as center point. The social input can be categorised into 4 stages, which are slight increment, huge increment, slight decrement and

huge decrement that can help to define how the system should choose the subsequent song across the song space.

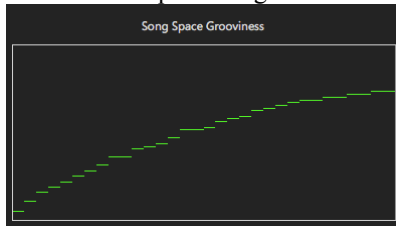


Figure 3. Social Input Panel

6.2 Smart Equaliser

Smart Equaliser creates a new approach for users to manipulate the equalization setting. There are 3 types of equaliser interface to suit different kinds of user.

Firstly, a set of parametric equalization is included for basic adjustment of low, mid and high frequencies.

Secondly, a typical 13 bands of graphic equaliser that is widely used in music players is prepared for users who prefer to have greater control yet it is simple to manipulate.

Thirdly, a fully customisable graphic equaliser gives more flexible adjustment for individual frequency bands. Users can draw a line across the space horizontally to increase or decrease the gain of specific bands. It is the most complex version but it also comes with the highest potential to discover new equalisation effect for audio tracks.

The algorithm is created by using convolution between the music signals with signals obtained from Fast Fourier Transform. The number of individual bands is 256 bands and equally divided from 20 Hz to 20kHz to give the maximum flexibility for the users yet it is not difficult to manipulate the music signals. Common equalisation presets are available for the ease of use while custom setting is also available for user to save as their personal preferences.

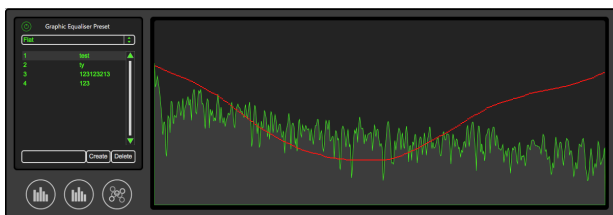


Figure 4. Smart Equaliser

6.3 Advance Setting Panel

Parameters for song space model, song selection and song transition are available for the system to determine a set of suitable songs for users.

The song space model can be varying based atmosphere mode which is targeting for the need of users while song recommendation is needed based on different scenario such as partying, house music for relaxing, chill and calm music before sleep and other possible scenarios. Different scenario setting generates the result of different space models based on different sets of feature combination.

Song selection automation features such as *Danceability*, *Grooviness* and proximity threshold help to customize the behaviour of system while generating playlist based on seed songs. Danceability is a high level descriptor that referred as the dance-ness of the song that can be defined as how significant the beat of the song that will make the listener feels like dancing. Grooviness is defined as the how the songs in the playlist build up or down the groove throughout the playtime.

7. SMART SONG TRANSITION

SmartDJ includes song transition features that it helps to transit from one song to another. The conventional DJ software focuses more on the manual features instead of automation. Therefore, we purposed several smart transition automation techniques inspired by the actual DJ skills for song transition. The transition comes in to amend where there is no perfect match for subsequence song by gradually pitching, adjusting BPM over the course of several minutes and creating equalisation mixing to prevent frequency band overlapping with each other.

7.1 Spectral Matching

Spectral matching is one of the basic DJ skills that apply the technique of tuning down the low frequency of current song while transiting to the next song. This technique helps avoiding the low frequency of both songs to overpower each other. This technique does not limit to only low frequency but mid and high frequency range as well. This techniques can simulate the actual DJ action which is turning up or down of the parametric equaliser.

EQ Blend [13] requires some equalization blending. The technique focuses on preventing the certain frequency bands of current song overpower the subsequence song. For an example, track A is an instrumental track with heavily concentrated on mid range of frequency spectrum while track B is a vocal track, mixing these 2 songs together will cause the frequency overpowering issue on mid range frequency band. DJs have to tune down the desired frequency band to make room for the next song to come in.

7.2 Beat Matching

Beat synchronization is another essential techniques that commonly used by DJs to do song transition. 3% rule applies where the tempo difference between 2 songs is within 3% when we change the tempo of the song. Two songs in F Minor that have BPM of 130 and 131 can be harmonically mixed together because the tempo difference is less than 3%. This helps to maintain the original pitch of the song without changing the speed significantly. Moreover, a purposed technique, "Modus Matching" helps to select the corresponding BPM for subsequence songs when there is not suitable songs fall within the 3% region. It selects the following song by twice or half of the tempo. [14]

7.3 Tonality Matching

The purpose of tonality matching is to ensure the transition can be done musically smooth. The concept of harmonic mixing is to ensure the song will be harmonically compatible based on certain conditions such as same key (Tonic), relative Major/Minor key, sub-dominant key (Perfect 4th) and dominant key (Perfect 5th). This involved in musical context where a song in Cm can be easily compatible with other songs with same key, its relative major D#/Eb, its sub-dominant Fm or its dominant Gm. [15]

A matrix of weights that representing the proximity between tonalities applies to select the subsequence songs with best harmonic matching possible. This matrix can be derived based on Camelot Wheel Chart [16] that shows the similarity between different keys.

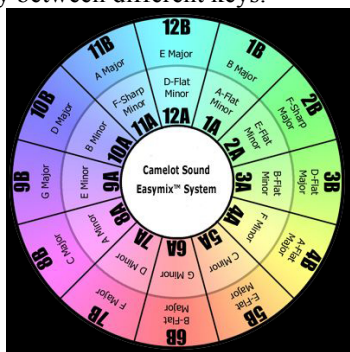


Figure 5. Camelot Wheel Chart

8. CONCLUSION

SmartDJ, other than a music player with DJ features, we proposed a novel way in which user can browse his/her music collection. With the implementation of the song space model, songs projected onto a Song Space Visualiser allows user to better understand the relationship between the music and to make a known song selection to better fit their current listening context. And with the DJ transition effects and automatic song selection, music listening will no longer be the same.

The objective of SmartDJ project is not just about developing a prototype but creating potential commercial product or at least it prepares ground for commercial application. Therefore, our project focuses on developing the front end (user interface) and back end (features extraction and dimension reduction) concurrently to create a new way of interacting with music. There are additional features can be included for further development such as song structure detection and user rating system to further enhance and complete the software functionality.

Acknowledgments

This project served as part of final year project for undergraduate studies of Information Engineering and Media at Nanyang Technological University, Singapore. PerMagnus Lindborg from School of Art, Media and Design, and Associate Professor Andy Khong Wai Hoong from School of Electrical and Electronics Engineering supervised this project.

9. REFERENCES

- [1] Schuller, Björn, Hage, Clemens, Schuller, Dagmar and Rigoll, Gerhard (2010) ‘Mister D.J., Cheer Me Up’: Musical and Textual Features for Automatic Mood Classification’, *Journal of New Music Research*, 39: 1, 13-34.
- [2] Watson, Diane & Mandryk, Regan L. (2012) “Modeling Musical Mood from Audio Features And Listening Context On An In-Situ Data Set”, University of Saskatchewan, 2012 International Society for Music Information Retrieval.
- [3] Pedro Davalos, “Automatic Music Genre Classification”, CPSC 633-600, Final Project Report, May 13, 2009.
- [4] Sam Clark, Danny Park, Adrien Guerard, “Music Genre Classification Using Machine Learning Techniques”, 2012.
- [5] Thibault Langlois, Gonçalo Marques, “A MUSIC CLASSIFICATION METHOD BASED ON TIMBRAL FEATURES”, 10th International Society for Music Information Retrieval Conference, ISMIR 2009
- [6] MATLAB Release 2010b. Natick, Massachusetts: © 1984-2010 The MathWorks, Inc.
- [7] Lartillot, Olivier, Toiviainen, Petri, “A Matlab Toolbox for Musical Feature Extraction From Audio”, International Conference on Digital Audio Effects, Bordeaux, 2007.
- [8] Max 6. © 2013 C’74. <http://cycling74.com>
- [9] Schwarz, Diemo, “ircamdescriptors~ real-time descriptor analysis for Max/MSP”, IRCAM Real-Time Music Interaction Team, France.
- [10] Arenas-Garcia, Jerónimo, Kaare Brandt Petersen, and Lars Kai Hansen. "Sparse kernel orthonormalized PLS for feature extraction in large data sets." *Advances in Neural Information Processing Systems* 19 (2007): 33-40
- [11] Schwarz, Diemo, “CataRT, Real-Time Corpus Based Concatenative Synthesis”, IMTR Team, IRCAM – Centre Pompidou, Paris, France, 2007.
- [12] Lillie, Anita Shen, “MusicBox: Navigating the space of your music”, Master Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2008
- [13] Cartledge, Chris, “EQ Mixing: Critical Techniques and Theory”, DJ TechTools, <http://www.djtechttools.com/2012/03/11/eq-critical-dj-techniques-theory/>, 11 March 2012.
- [14] “How to: Understanding key and tempo in harmonic matching”, Harmonic Mixing Community, <http://community.mixedinkey.com/Topics/1767/how-to-understanding-key-and-tempo-in-harmonic-mixing>, October 2008.
- [15] “Mixing Harmonically”, MixShare: ReWiki, http://www.mixshare.com/wiki/doku.php?id=mixing_harmonically, 9 June 2011.
- [16] Davis, Mark, “Harmonic Key Selection”, Camelot Sound, <http://www.camelotsound.com/Easymix.aspx>.