Classification of modern pop songs into ‘Day’ and ‘Night’ songs.

Raunaq Aman Jaswal
Department of ECE
rjaswal@ur.rochester.edu

ABSTRACT

There are songs that are evocative of and/or better suited to a particular time of the day, either done intentionally by the composers of the song, and it may contain latent features, in its instrumentation and/or lyrics that makes it suitable for that time of the day. This paper attempts to classify audio tracks into ‘Day’ or ‘Night’ songs based on the features of its musical instrumentation. Timbral, temporal, spectral, perceptual as well as rhythmic features of the audio tracks are extracted and then fed to an SVM classifier, classifying ‘Night’ songs from the ‘Day’ songs. Expansion and future work of the paper will address classification of songs to a particular time of day, and improvement of feature selection and the classification method.

1. INTRODUCTION

Much of the songs sound suited to a particular time of the day. Many music-service providers like Spotify and Apple Music curate playlists which is based on a particular time of a day, and many of the songs in a particular playlist share similar qualities. ‘Morning’ and ‘Day’ songs tend to be more ‘acoustic’ in its instrumentation and the vocals seem to be cleaner as compared to the ‘Evening’ and ‘Night’ songs, where the instrumentation seems much more ‘electronic’, ‘smokier’ and filled with ‘echo’ and ‘reverb’ and the vocals seem to be dripped in effects. The tempo of a ‘Night’ song also feels slower in comparison to an ‘Evening’ song.

Even in the earlier days, composers used to create pieces which was inspired by, or evocative of, the night, called ‘Nocturne’ These included several instrumentations and passages which were intended to be played at night. Brian Eno, an electronic producer has composed several albums worth of material, which is meant to be played at late nights.

Previous research has tried extract higher-level semantic features from songs, such as the mood in a song or the energy of the performance in a song. However, none of the research has been done on features of a song that evokes or is reminiscent of a particular time of day or point of time in year, for example, a summer album.

The main motivation is to find and extract which features in its instrumentation makes a song a ‘Day’ or ‘Night’ song in a more objective sense. Based on the difference in the instrumentation and vocals of the song, the aim of the project is to classify songs into, ‘Day’ or ‘Night’ songs. The snippet of a song present in the database should be representative of the whole song, that is, it should contain elements, timbral as well as rhythmic that are sustained.

Earlier, the paper planned to extract features based on the “electronicness” and “reverbiness” of an instrument, that is how much has the instrument been processed, in contrast to an acoustic instrument in which only minimal processing has taken place as well as how reverb-heavy the vocals were. However, due to lack of published research done in that area, as well as due to time constraints, the paper will not go towards those areas, instead using and extracting more conventional timbral and rhythmic features.

First, the relevant timbral, temporal, spectral, perceptual as well as rhythmic features from the audio source are selected and then calculated. Next, the method uses Principal Component Analysis to filter out nonseparable or noisy features and reduce the feature vectors’ dimensionality. then the reduced feature set are sent for classification.

By finding out and selecting the optimal set of features that best describes the difference between the ‘Day’ and ‘Night’ songs in terms of its instrumentation and vocals, it can be used for content-based filtering, with regards to music recommendation. These features can also be used to algorithmically generate songs which is best suited to the time of the day, when music generation has evolved to the point where it can self-compose various kinds of music.

2. METHOD

2.1 Dataset

The dataset used is the GTZAN Dataset for the Music Analysis, Retrieval and Synthesis for Auido Signals project. This dataset has been used for various audio-genre based classification problems, and since this project uses similar techniques, selection of the dataset seemed pertinent.

The dataset consists of 1000 audio tracks each 30 second long. The audio tracks have a sample rate of 22050 Hz and are 16-bit monochannel .wav files. The original dataset consisted of 10 labels corresponding to 10 major genres which includes 100 songs each.

There were no datasets which labelled songs in accordance to the aim of the project, and the only place these labels were available were in curated playlists on music streaming services.

Thus, the dataset was thus labelled personally, which can potentially lead to bias in the labelling.

The training/testing split was done on 900 of the songs, discarding classical songs, and the split was 80/20.

2.2 Preprocessing

The amplitude of each audio track was normalized and converted to the following representations for further feature extraction. [1-3][8][9]

1) Short-Time Fourier Transform: To calculate instantaneous descriptors.

2) Mel-Scale: To calculate the Mel Frequency-Cepstrum Coefficients

3) Bark Scale: To calculate perceptual descriptors

Prior to calculating the bark spectrogram, a filter simulating auditory response of the mid-ear was applied to the magnitude spectrogram.

2.3 Feature extraction and reduction

2.3.1 Feature Extraction

Some of the features extracted from the audio signal representation are given in Table 1:

Representation	Features
Spectral Features	• Spectral Centroid, • Spectral Flux, • Spectral Flatness, • Spectral Roll-off, • Spectral Kurtosis, • Spectral Crest
Mel-frequency Features	• 13 MFCCs
Perceptual Features	• Sharpness • Spread

Table 1. List of some of the features selected and their representation

The following features are calculated from the calculated representations for each frame and then averaged out throughout all frames(m(f) is the magnitude of the FFT at frequency bin f and N the number of frequency bins):

• Centroid:

$$C = \frac{\sum_1^N fm(f)}{\sum_1^N m(f))} \tag{1}$$

The Centroid is a measure of spectral brightness.

• Rolloff: is the value R such that:

$$\sum_1^R m(f) = 0.85 \sum_1^N m(f) \tag{2}$$

The rolloff is a measure of spectral shape.

• Flux:

$$F = \|m(f) - m_p(f)\| \tag{3}$$

Where $m_p(f)$ denotes the FFT magnitude of the previous frame in time. Both magnitude vectors are normalized in energy. Flux is a measure of spectral change.

• Spectral Flatness:

$$SF = \frac{(\prod_{k=num\_band} m(k))^{1/K}}{\frac{1}{K}\sum_{k=num\_band} m(k)} \tag{4}$$

where m(k) is the amplitude in frequency band k, and K is the total number of frequency band. Spectral Flatness is a measure o the noisiness of the spectrum.

• Spectral kurtosis:

$$Kurtosis = \frac{\sum_1^N (f - mean)^4 m(f)}{\sum_1^N m(f)\sigma^4} \tag{4}$$

Kurtosis indicates the peakedness of the distribution.

• Sharpness:

$$A = 0.11 \frac{\sum_{z=1}^{nband} z.g(z).N'(z)}{N} \tag{5}$$

Where z is the Bark band number, g(z) is the band number function and N'(z) is the specific loudness, that is the loudness associated to each Bark Band.

• Spread:

$$ET = \left(\frac{N - max_z N'(z)}{N}\right)^2 \tag{6}$$

Where N is the total loudness of the frame.

Although, recent automatic music genre classification calculated the features by splitting audio tracks using onset-based detection [4], the features in this case were calculated throughout the audio snippet. The reason for this is the audio snippet should contain features that are prevalent throughout the track and not just in certain musical moments.

2.3.2 Feature Reduction:

After the extraction the features, they are then concatenated to form a feature vector. After that, principal component analysis (PCA) is used to reduce the feature vectors' dimensionality [5][6]. Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert possibly related features into a set of values of uncorrelated features called principal components. This will ensure the noisy or nonseparable features are filtered out and the classifier receives varied data.

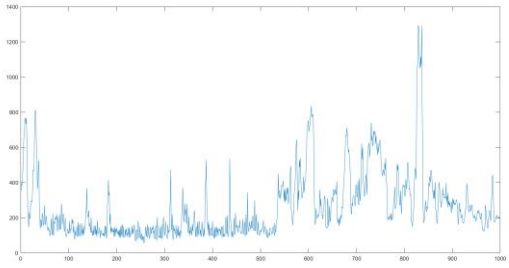

Figure 1. Spectral Centroid variation in a 'Night' song.

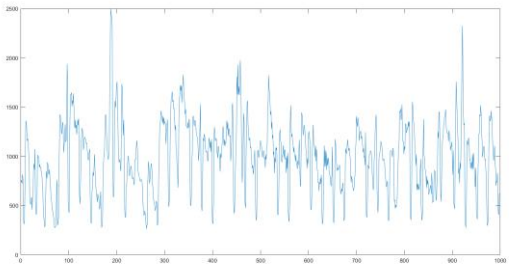

Figure 2. Spectral Centroid variation in a 'Day' song.

2.4 Classification

Since right now it is a binary and a supervised classification problem, support vector machine (SVM), which is a supervised machine learning algorithm which can be used for both classification or regression challenges, would be adequate. Further on, with more classification labels, a more appropriate and a better classifier may be needed. The kernel used for the classifier was a Gaussian Kernel.

3. RESULTS

The testing was done using MATLAB. Future works would include producing the code in a deployable environment.

After limited testing, the confusion matrix of the best classification result is depicted in Table 2:

	Predicted Day	Predicted Night
True Day	62.98%	37.02%
True Night	42.35%	57.65%

Table 2. Confusion matrix of the output of the test

As seen above, the results don't seem too promising using conventional classification features. Furthermore, two of the features extracted explained at least 95% of the variability between the labelled songs.

4. CHALLENGES

Earlier direction of project involved extracting more latent and objective features. Despite attempts to objectify features that make a song 'Day' or 'Night', it is still a very subjective opinion. Song composers try to make songs which was inspired by, or evocative of, the night. However, that may not translate well to the listener. General features, used for other classification problems, such as genre classification and mood classification may not work very well and thus may not be sufficient when approaching this particular problem. Some songs may be evocative of or suited to both night and day. There may be some weakness in the assumption made in **1.** that songs suited to the night necessarily have those features, although it is assumption made by scouring the playlists curated by music streaming services such as Spotify and Apple Music.

5. CONCLUSION AND FUTURE WORK

Future work could involve extracting features which are better representative of the differences between these songs. Although the results haven't been too promising, improvements in feature selection and classifiers may help in reducing the errors. Also, figuring out the instrumentation and what type of audio effect feature is there in the instrumentations and vocals in polyphonic music may also be considered as a future research. Also, the features which are strongest could be used in algorithmic/ computer music composition to create pieces that similarly evoke feelings of night music.

6. REFERENCES

[1] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, 2004.

[2] Emmanouil Theofanis Chourdakis and Joshua D. Reiss: "Automatic Control of a Digital Reverberation Effect using Hybrid Models," *AES 60TH INTERNATIONAL CONFERENCE, Leuven, Belgium, 2016 February 3–5*

[3] T. Ganchev, N. Fakotakis, and G. Kokkinakis. "Comparative evaluation of various MFCC implementations on the speaker verification task." *Proceedings of the SPECOM-2005, 2005*

[4] K. West and S. Cox. "Finding an optimal segmentation for audio genre classification,"*6th International Conference on Music Information Retrieval*, pages 680–685,2005.

[5] Y. Lu, I. Cohen, S. Zhou, X, and Q. Tian. "Feature selection using principal component analysis,"

Systems Science, Engineering Design and Manufacturing Informatization (ICSEM), volume 1, pages 27–30. IEEE, 2007

[6] Horn, J. L. (1965). *"A Rationale and Test for the Number of Factors in Factor Analysis,"* Psychometrika, 32, 179-185.

[7] Michael Stein, "Automatic Detection of Multiple, Cascaded Audio Effects In Guitar Recordings," *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria, September 6-10, 2010.*

[8] G. Tzanetakis P. Cook, "Automatic Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing, 2002*