Vol.2011-MUS-89 No.9
2011/2/11

Segmentation of Music Using Physiological Data

Rafael Cabredo,[†1] Roberto Legaspi[†1]
and Masayuki Numao[†1]

Automated segmentation of music is currently performed using music features related to rhythm, timbre and harmony. A method for segmenting music by clustering physiological data is presented in this paper. Results are compared against manual segmentation of a song list of 24 songs. Standard classes to clusters evaluation method is used to determine accuracy of clustering. Best results from the experiments use features related to blood volume pulse (BVP).

1. Introduction

Music has a high level structure created by either a repetition of a sequence of features or a change in a constant feature. This paper investigates a way of dividing a song into its high level sections such as introduction, or verse using physiological readings of the music's listener and chord progressions of the music. We are interested in learning how these segments and its musical properties constitute emotional changes in its listeners. The relation of music, emotion and physiological responses has been studied in 4). While listening to music, listeners react to certain parts of the music. We hypothesize that changes in physiological responses are heightened when the music moves from one segment to the next.

Similar research work on automatically segmenting music into high level musical structure has been done in[3),5)]. It is a fundamental problem in computational music theory and has various applications, such as used in music information retrieval, copyright infringement resolution, music navigation, and finding repeating structures in music.

This paper is organized as follows. First, the methodology we employed for collecting data is explained, followed by the data preprocessing details. Next, the machine learning task is described and the results and observations of the experiments. Finally, a conclusion and future work is given.

†1 The Institute of Scientific and Industrial Research, Osaka University

2. Methodology

Our approach requires collecting psycho-physiological data from a subject while he listens to music. For the research, we concentrate on analysing data from one subject (a 22-year male graduate student). The data was recorded using BioGraph Infinity System.[⋆1] Three sensors were used to record data on blood volume pulse (BVP), skin conductance (SC) and respiration rate (RR) separately.

2.1 Music selection

The songs used for the research are part of the music dataset described in 8). Songs were selected based on 3 constraints. First, the song should not have any key and tempo changes. Second, the song should have complete chord and segment annotations. Last, the song is in a major key. Using this criteria, 83 songs were selected which include 77 songs from the Beatles, 4 Queen songs, and 2 Carole King songs. Tempo and key information of the music data set is shown in **Table 1**. The chord and segment annotations from the isophonics dataset include onset and offset times for each chord and segment change.

Table 1 Summary of music

Tempo	Key											Total
	B	**E**	**A**	**D**	**G**	**C**	**F**	**B♭**	**E♭**	**A♭**	**F♯**	
Larghetto			1									1
Adagio		2	3	1	2	2			1			11
Andante	1	2	4	2	3	2	1	1		1		17
Moderato	1	5	4	1	2	1	2					16
Allegro		4	6	3	3	4	2	1			1	24
Presto	1	4	1	4	2	1	1					14
Total	3	17	19	11	12	10	6	2	1	1	1	**83**

Larghetto : 60–66bpm[⋆2] Adagio : 66–76bpm Andante : 76–108bpm
Moderato : 108–120bpm Allegro : 120–168bpm Presto : 168–200bpm

⋆1 About BioGraph Infinity System. Thought Technology Ltd. 8 Dec 2010. http://www.thoughttechnology.com
⋆2 bpm:beats per minute

Our subject listened to 83 songs via audio-technica closed headphones (ATH-T400) connected to a computer in a controlled experiment room. Several sessions were needed to allow the subject to listen to all the songs without making the subject feel stressed. Each session took about 20 minutes allowing the subject to listen to 7 to 9 songs per session. One week was needed to complete the data collection. Sessions were held at the same time of the day throughout the week.

Before each session ended, the subject also annotated the songs listened to for the session. The subject was instructed to rate the music according to how he enjoyed listening to it. This rating was used to determine the songs to be included in the machine learning task.

2.2 Feature set

From the data collected, feature vectors were constructed for each song. Each vector consists of 45 attribute values : 43 attributes from physiological data, a chord label and the segment label, which was used for verification. **Table 2** summarizes the set of features from physiological data. The BioGraph Infinity System includes software that automatically computes from the acquired raw signals cepstral attribute values within a specified epoch as indicated by a sliding time-slice window. The chord and segment labels were provided by human annotators as described in 8).

Table 2 Physiological features

Sensor	Features
BVP (36 features)	• BVP: raw signal, amplitude mean(μ), peak freq μ (Hz), LF/HF (μs), LF/HF (epoch μs) • inter-beat interval(IBI): peak freq, std(σ)/SDDR, epoch σ, peak amplitude, peak amplitude max, NN interval • HR from IBI: HR μ, HR σ, HRMAX-HRMIN, HR epoch μ HRMAX-HRMIN μ (b/min) • HR standard freq bands (VLF,LF,HF): %power, total power,%power μ, total power μ, %power, epoch μ, total power epoch μ
RR (4)	• raw signal, rate, rate μ (br/min), rate epoch μ
SC (3)	• raw signal, μ (microS), epoch μ

2.3 Data set

The data set is comprised of the physiological readings and music information for selected songs. Based on the annotations provided by the subject, it was found that not all songs provided the same level of enjoyment. Thus, only highly rated songs were included for the experiment. Twenty-four songs were selected from the data collection. These songs were selected from 3 different tempo groups: adagio, allegro, and presto. For every song, 4 data sets were constructed – one for every physiological sensor and one data set that combines all data. In total, 96 data sets were prepared for the automated segmentation task.

Each data set was also annotated with the chords and segment labels. All feature vectors representing the physiological signals at the time t are labeled with the current chord and segment being playing at t. All labeled feature vectors were normalized to the range [0,1] since the components differ in the scales in which their values lie. Equation (1) was used for normalizing the values.

$$A_{normalized} = \frac{A - min(A)}{max(A) - min(A)} \tag{1}$$

3. Machine Learning Task

To identify the music segments, k-means clustering[6),7)] is used. Given a set of observations $(x_1, x_2, ..., x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition n observations into k sets ($k \leq n$)$S = \{S_1, S_2, ...S_k\}$ so as to minimize the within-cluster sum of squares :

$$\underset{x}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_j \in S_i^{(t)}} \|x_j - \mu_i\|^2 \tag{2}$$

where μ_i is the mean of points in S_i.

The algorithm for clustering is described as follows:

Given an initial random set of k means $m_1{}^{(1)}, ..., m_k{}^{(1)}$, the algorithm proceeds by alternating between two steps:

(1) Assignment step: Each observation to the cluster with the closest mean is assigned.

$$S_i^{(t)} = \left\{ x_j : \left\| x_j - m_i^{(t)} \right\| \leq \left\| x_j - m_{i^*}^{(t)} \right\| \forall i^* = 1, ..., k \right\} \tag{3}$$

(2) Update step: Calculate the new means to be the centroid of the observations in the cluster.

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j \tag{4}$$

The algorithm is said to have converged when the assignments no longer change.

3.1 Simple k-Means

The simple k-means implementation of WEKA[1)] was used for clustering the different datasets. This implementation handles a mixture of categorical and numerical attributes. The algorithm uses Euclidean distance measure to compute distances between instances and clusters.

Different values of k was used to see how clusters are formed as k increases.

3.2 Feature Selection

Feature selection is the process of identifying and removing as much of the irrelevant and redundant information as possible. Reducing the number of features help improve concept of generalization and reduce computational costs. A search strategy is also needed to explore the space of all possible features. In our work, correlation-based feature (CBF) subset selection[2)] using best first search strategy was used. This algorithm evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them.

4. Results

Ideal segmentation of the music is indicated when instances that belong to the same segment are grouped in the same cluster. The average results using k-means on the data sets for every sensor type is shown in **Fig. 1**. The best average result of 60.92% was obtained by combining all physiological data of the data set and using $k = 10$. However, accuracy using BVP data alone is almost as accurate.

Figure 2 shows an example of how the data set of the song "So Far Away" by the Beatles was clustered. Examining the results across the data, it is observed that shorter segments (i.e., those labeled as INTRO, OUTRO, BREAK) are more likely to have its instances clustered together than longer segments. Cluster assignments contain a mixture of instances from different segments of the song.

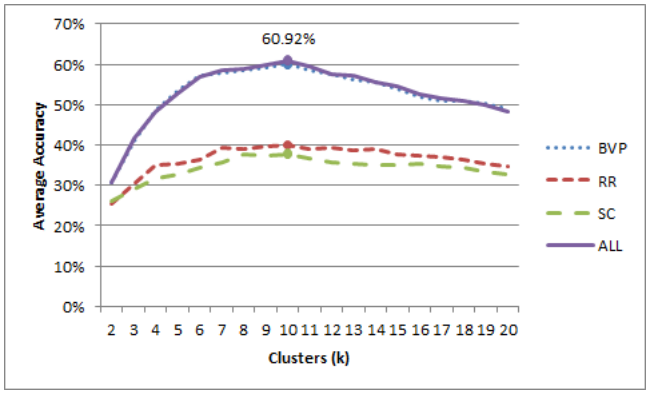

Fig. 1 Clustering accuracy using K-Means

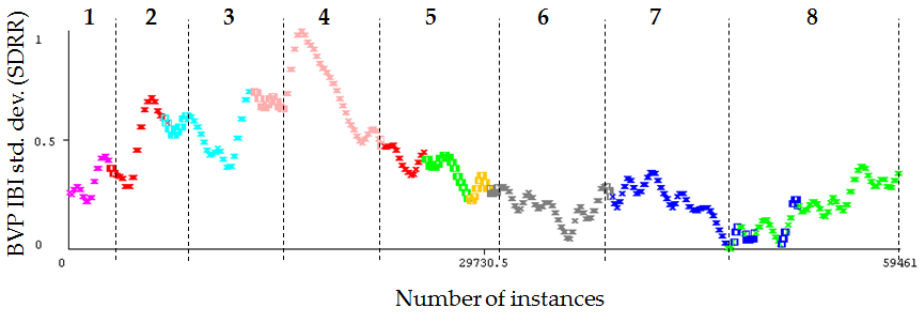

Fig. 2 Visualization of cluster assignments for "So Far Away." Colors represent different clusters and horizontal divisions indicate the music segments

Typically, segments that are adjacent to one another have instances that are placed in the same clusters. This can be interpreted as adjacent segments have similar physiological signals. The data also shows that there are segments wherein the subject experienced strong emotional responses evidenced by dominant peaks in BVP data.

Since using BVP data alone was observed to be sufficient to cluster instances, we applied feature selection on those data sets. **Figure 3** shows the improvements after using feature selection. From 36 features, CBF subset selection reduced the feature set between 4 to 9 features. Average performance after feature selection

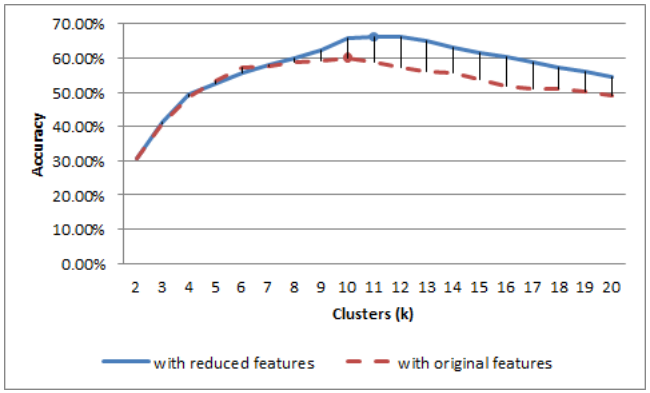

Fig. 3 Clustering accuracy after feature selection

Fig. 4 Clustering accuracy using respiration rate grouped by tempo

was 65.69% using $k = 10$ and 66.22% using $k = 11$.

Although respiration rate has a low accuracy rating for segmenting music we observe that its average accuracy is directly proportional to the increase of tempo of data sets as shown in **Fig. 4**.

5. Conclusion and Future Work

We observe from the preliminary work that a person listening to music manifests physiological reactions that can be used for analyzing music. In this study, BVP-related features were identified to be the most promising to use for identifying music segments. Further research on the music features of each music segment is needed to understand how these features affect people listening to music. The current clustering technique and similarity measure does not consider the musical structure of chords. Development of other techniques that considers these information are needed to improve accuracy of music segmentation.

References

1) M.Hall, E.Frank, G.Holmes, B.Pfahringer, P.Reutemann, and I.H. Witten. The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November 2009.
2) M.A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, 1998.
3) K.Jensen. Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP Journal on Advances in Signal Processing*, 2007:1–12, 2007.
4) J.Kim and E.Ande. Emotion recognition based on physiological changes in music listening. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(12):2067 –2083, 2008.
5) M.Levy, K.Noland, and M.Sandler. A comparison of timbral and harmonic music segmentation algorithms. In *ICASSP 2007. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume4, 2007.
6) S.P. Lloyd. Least square quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
7) J.MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume1, pages 281–297. Univ. California Press, Berkeley, Calif., 1967.
8) M.Mauch, C.Cannam, M.Davies, C.Harte, S.Kolozali, D.Tidhar, and M.Sandler. OMRAS2 metadata project 2009. In *10th International Conference on Music Information Retrieval Late-Breaking Session*. Kobe, Japan, 2009.