VOL. 10, NO. 5, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

ANALYSIS OF MEL BASED FEATURES FOR AUDIO RETRIEVAL

R. Christopher Praveen Kumar and S. Suguna
Department of Electronics and Communication Engineering, V.S.B. College of Engineering, Technical Campus, Coimbatore, India
E-Mail: Chris2praveen@gmail.com

ABSTRACT

Nowadays the electronic gadgets have been updated to store large amount of music information. It is necessary to have an efficient retrieval system to choose the required data. The important task in audio retrieval system is feature extraction. In the feature extraction stage, the feature which gives relevant information about music has to be extracted. In this paper, various Mel based feature which includes Mel Frequency Cepstral Coefficient (MFCC), Delta MFCC (DMFCC), Double Delta MFCC (DDMFCC), hybrid feature (MFCC+DMFCC+DDMFCC) has been analyzed for audio retrieval system. It has been found out that the audio retrieval system which makes use of hybrid feature will provide better result compared to the other features.

Keywords: audio retrieval, hybrid feature, MFCC, feature extraction.

1. INTRODUCTION

In modern world most of the people relax themselves by listening to music. There is in need of a storage device which can store large amount of music information [1]. To match up with people's expectation new electronic gadgets has come into existence for storing the large collection of music file. In order to select the particular audio file from the database, an efficient audio retrieval system is required. Audio retrieval deals with retrieval of similar pieces of music, instruments, artists, musical genres, and the analysis of musical structures [2]. Among the various stages in the audio retrieval process, feature extraction stage is very important stage. The feature extraction process involves extracting the necessary feature from the audio file which will give sufficient information about the audio file [3]. Based on the information obtained from the feature extraction process, the audio files can be discriminated among the other audio files. From the literature survey, it has been found out that timbral feature will provide meaningful information about the audio file. The timbral feature can be represented using MFCC. In this paper the analysis has been done on various Mel based feature which includes derivatives of MFCC feature for the audio retrieval task. The audio retrieval task has been explained in the latter sections and the audio retrieval performance of various Mel based features has been discussed. Among the various Mel based feature, audio retrieval system which makes use of hybrid feature provides better audio retrieval results.

2. RELATED WORKS

The previous works related to the feature extraction process in audio retrieval task has been discussed in this section. In [4] the importance of Mel Frequency Cepstral Coefficient (MFCC) feature extraction has been discussed and text dependent speaker identification system has been designed using Mel Frequency Cepstral Coefficient (MFCC) feature. The timbral based approach for audio classification based on the Gaussian mixture model and clustering technique has been discussed in [5].The audio classification method which makes use of support vector machine and radial basis neural network has been proposed in [6]. The feature used in the audio classification task is MFCC and linear predictive coefficient. In [7] audio tag annotation which was performed on CAL500 and CAL10k corpora makes use of Dirichlet mixture model (DMM) has been discussed. A novel approach [8] which make use of ensemble classifier for automatic annotate and audio retrieval has been performed on a database which consist of 2,473 clips and the duration of each clip is 10 seconds or less. In [9], content based audio retrieval based on Distance-from-boundary (DFB) and classification task based on support vector machine (SVM) has been performed on a database which consists of 409 sounds of 16 classes. A novel approach has been proposed in [10] for classifying General Audio Data (GAD). The features used for classification task are Mel Frequency Cepstral coefficient (MFCC) and Linear Predictive Coefficient (LPC). In order to minimize the noise occur at the boundary of audio segments during classification, segmentation pooling method has been used. Currently research is going on towards effective classification of musical instruments. In [11], timbral model has been designed with the help of a mixture of Gaussian distribution for a space of Cepstral coefficient. Automatic audio classification task based on statistical pattern recognition classifiers has been proposed in [12]. The feature sets used for the classification task can be used to represent timbral texture, rhythmic content and pitch. In [13], for both voice print and dynamic characteristics of the spoken digits an improved technique of feature extraction using Weighted MFCC is used. Improved Features for DTW (IFDTW) has been used for feature recognition. From the above discussion, it was found that MFCC feature provide better result for audio retrieval and classification task. The steps involved in hydrid feature extraction process have been discussed in the following sections.

3. METHODOLOGY

a) Audio retrieval

Audio retrieval involves retrieving the similar piece of audio file. The important process in the audio retrieval process involves feature extraction. The feature indicates numerical representation of audio file because it is difficult to process the raw audio file. The feature has been extracted from each audio file in the database and stores it in feature database. The feature has been extracted from the query audio file and it has to be compared with each audio file feature in the feature database.

When query audio file feature matches with the feature database, then the corresponding audio file will be retrieved.

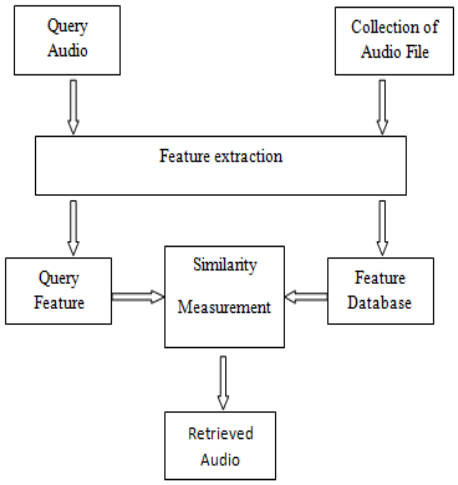

Figure-1. General block diagram for audio retrieval.

In this work, the similarity measurement between audio database feature and query audio file feature has been done with the help of Euclidean distance. In the Euclidean distance measurement, the query audio file is compared with each audio file in the audio database. For the similar audio file, the distance measured is small, while for the different audio file the distance is large. In this manner, the similarity between the audio file can be identified using Euclidean measurement. The formula used for this measurement is as follows

$$d(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \quad (1)$$

b) MFCC and its derivatives

MFCC is the most widely feature in audio processing. It indicates the cepstral representation of audio file. MFCC involves the collection of Mel Frequency Cepstrum (MFC) which is the representation of short term power spectrum of audio file. The audio signal is not constant, it keeps on changing. The changes in the audio signal lead to MFCC derivatives.

c) Hybrid feature extraction

The hybrid feature includes the combination of MFCC, Delta MFCC and Double delta feature. The hybrid feature extraction process involves several stages. The stage by stage process has been shown in Fig.2.Initially the audio signal

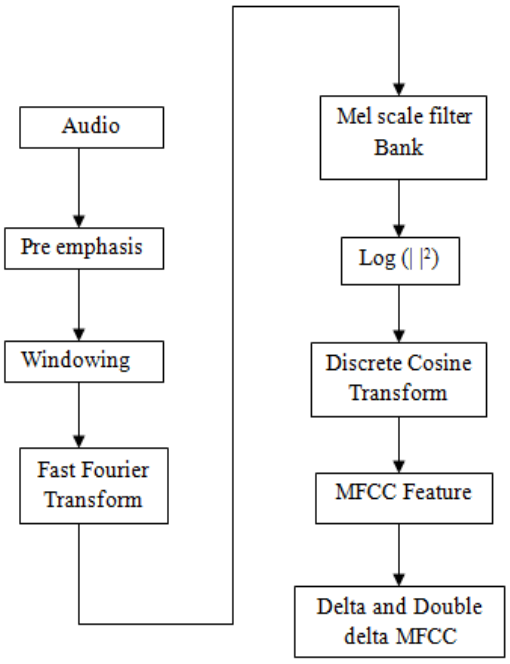

Figure-2. Hydrid Feature extraction process.

Shown in Figure-3 is applied to pre emphasis stage which is the first stage in the hybrid feature extraction process. Pre-emphasis involve boosting the high frequency components in the audio signal.

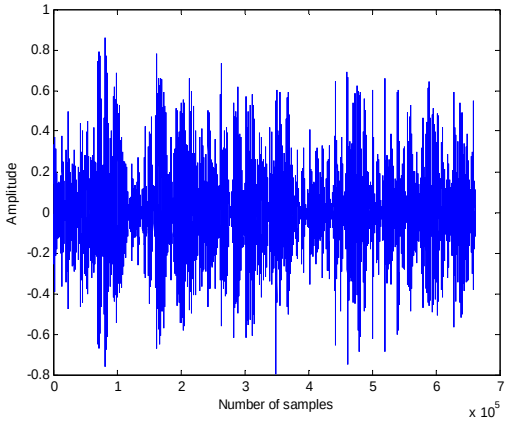

Figure-3. Audio signal.

By boosting the high frequency components of the audio signal, the additional information can be gathered from the audio signals.

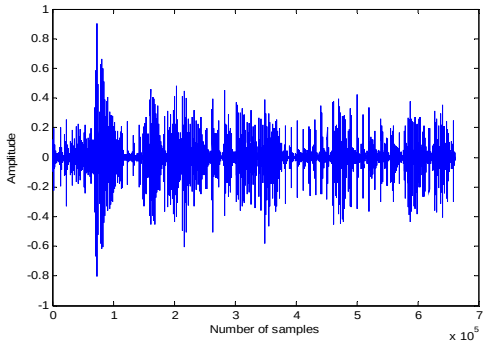

Figure-4. Pre-Emphasized signal.

As audio signal is not constant one, it is necessary to obtain information from small enough region. Hence the entire audio signal has to be divided into small frames of size 10-25ms and for each small segment hamming window has to be applied. This stage will reduce the discontinuity near the boundary of the audio signal.
The hamming window is given by the equation [2]

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{L-1}\right) & 0 \le n \le L-1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

In order to analyze the variation in each frame of the audio signal, it is better to analyze the audio signal in frequency domain. Hence Fourier transform has been applied to each frame of the audio signal. In Mel filter bank, triangular band pass filter is used. The triangular band pass filter has been applied to the frequency response of each frame of the audio signal. The Mel curve obtained in this process is shown in Figure-5. This process will reduce the feature size and smoothens the magnitude spectrum thereby harmonics gets flattened.

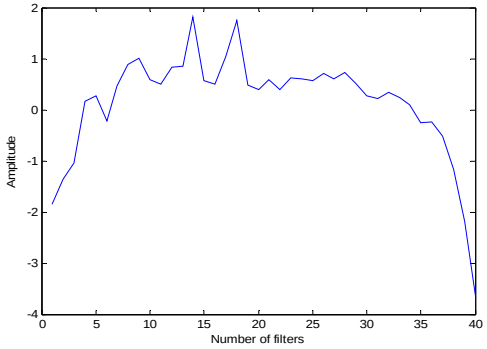

Figure-5. Mel Frequency curve.

The logarithmic computation has been done to reduce the dynamic range of the value. Next stage involves applying discrete cosine transform to the Mel spectral feature in order to de-correlate the Mel feature. The expression used to calculate the MFCC feature is given by [3]

$$c_n = \sum_{k=1}^{K} \log(x(k)) \cos\left[(k-0.5)\, n\pi / K\right] \quad (3)$$

The delta coefficient d_i can be computed at frame 'i' in terms of basic MFCC coefficient by the following expression

$$d_i = \frac{\sum_{n=1}^{N} n(c_{n+i} - c_{n-i})}{2\sum_{n=1}^{N} n^2} \quad (4)$$

The double delta MFCC feature can be calculated using above equation by replacing MFCC coefficient with delta coefficient. The combination of the above mentioned features forms hydrid feature which is shown in Figure-6.

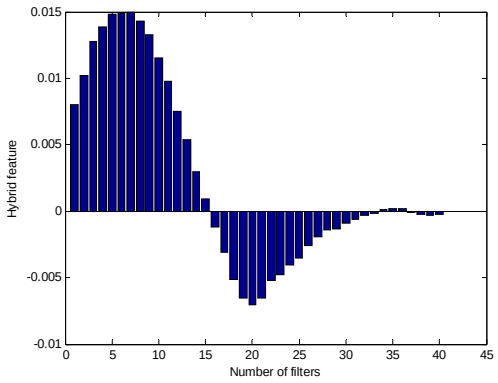

Figure-6. Hybrid feature.

d) Performance measure

The performance of the retrieval system can be analyzed by measuring the precision and recall values. The precision and recall values will determine the quality of audio retrieval system. The precision value indicates the relevant audio file from the retrieved files while recall value indicates retrieving relevant audio files. The formula for calculating precision and recall value are as follows

$$\text{Precision} = \frac{\text{Number of relevant audio}}{\text{Number of retrieved audio}} \quad (5)$$

$$\text{Recall} = \frac{\text{Number of relevant audio}}{\text{Total number of audio file}} \quad (6)$$

4. RESULTS AND DISCUSSIONS

In this section the audio retrieval performance using Mel based feature has been discussed. The audio

retrieval task has performed on the MATLAB software. The audio retrieval task make use of GTZAN database which consist of 10 genre namely blue, classical, country, rock, hip-hop, disco, Jazz, pop, Metal, Reggae where each genre consist of 100 songs. The Mel based features has been used to find audio similarity among the audio files. With the help of various Mel based feature, audio retrieval task has been performed and its performance has been discussed for different feature using precision and recall value.

Based on the query audio file, relevant audio file has been identified from the retrieved audio file using various Mel based feature and it has been found out that Hybrid feature provides better precision value compared to other Mel based feature. The mean precision rate for each Mel based feature for top-*K* audio file has been shown inFigure-7.

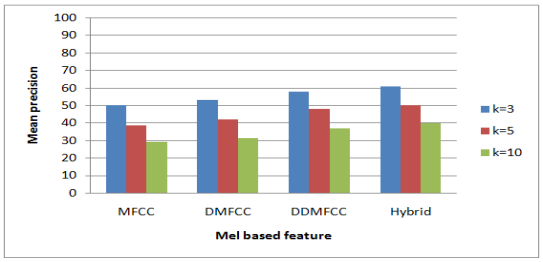

Figure-7. Mean precision for top-*K* (K=3, 5, 10) audio file

The important task in the audio retrieval process is feature extraction. The time required to extract each feature is important. Here the time taken to extract various Mel based feature has been discussed and it has been found that hydrid feature require more extraction time as it is combined feature and also the difference in extraction time among other feature is appreciable one.

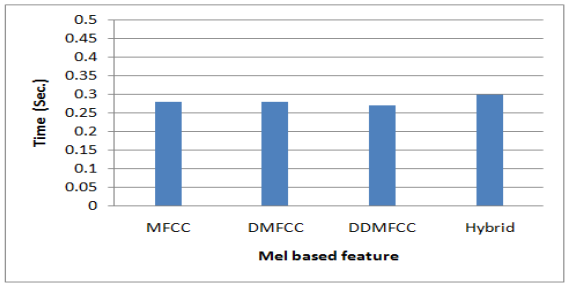

Figure-8. Feature extraction time.

After the feature extraction process the similarity measurement has to be done. In this stage the query audio file has been compared with the each audio file in the database. The similarity measurement time for Mel based feature has been shown in Figure-9 and it was found out that hybrid feature requires more time compared with other feature but it is appreciable one.

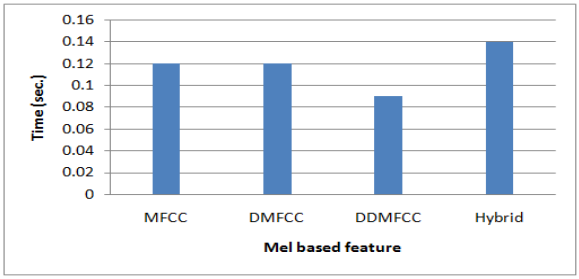

Figure-9. Similarity measurement time.

It is necessary for the audio retrieval system to provide relevant audio file based on the query. The relevant audio file retrieved using Mel based feature in the audio retrieval system has been calculated and it was found out that hybrid feature provides better recall value compared to other feature which shows that hybrid feature provide better recall value. The recall value plotted in the chart is shown in Figure-10.

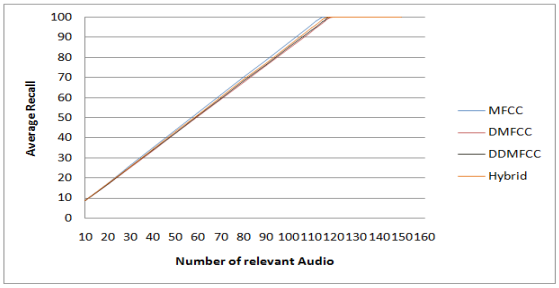

Figure-10. Average recall value.

As it was found out that hydrid feature provide better results compared to other Mel based feature in the audio retrieval task. The relevant audio file retrieved for various genres using hydrid feature has been found and tabulated in Table-1.

Table-1. Retrieved relevant audio file (Top-5).

Query	Top 5 Relevant Audio retrieved				
Blues00	Blues00	Blues19	Hip-hop26	Rock28	Pop34
Classical00	Classical00	Jazz11	Country99	Pop42	Rock47
Country00	Country00	Pop51	Hip-hop70	Jazz34	Disco28
Disco00	Disco00	Classical99	Rock36	Rock84	Metal50
Hip-hop00	Hip-hop00	Metal08	Blues28	Pop64	Disco31
Jazz00	Jazz00	Hip-hop40	Jazz07	Rock10	Country22
Metal.00	Metal00	Hip-hop01	Rock69	Hip-hop14	Metal26
Pop00	Pop00	Country20	Blues08	Hip-hop19	Jazz47
Reggae00	Reggae00	Classical22	Metal34	Metal94	Classical83
Rock00	Rock00	Blues07	Metal37	Disco61	Classical03

5. CONCLUSION AND FUTURE WORK

In this paper, the audio retrieval process has been discussed. The significance of feature extraction process in the audio retrieval task has been highlighted. Based on the literature survey, it was found that timbre feature provide meaningful information about audio signal. In order to represent the timbral feature, MFCC feature has been used. In this paper, various Mel based feature has been analyzed for audio retrieval process. The feature extraction process of hybrid feature has been explained. The audio retrieval process which makes use of Mel based feature has been performed on GTZAN database and it was found that among various Mel based feature hybrid feature provide better precision and recall value. In the future along with Mel based feature, other features will be considered for audio retrieval task.

REFERENCES

[1] Tomi Kinnunen. and Rahim Saeidi. 2012. Low-Variance Multitaper MFCC Features: A Case Study in Robust Speaker Verification. IEEE Transactions on Speech, Audio and Language Processing.

[2] Masayuki Suzuki. And Takuya Yoshioka. 2012. MFCC Enhancement Using Joint Corrupted and Noise Feature Space for Highly Non-Stationary Noise Environments. ICASSP.

[3] Dalibor Mitrovi´C., Matthias Zeppelzauer. and Christian Breiteneder. 2010. Features For Content-Based Audio Retrieval. Advances in Computers. Vol. 78, pp. 71-150.

[4] Vibha Tiwari. 2009. MFCC and its applications in speaker recognition. International Journal on Emerging Technologies.

[5] Thibault Langlois. and Gonc Alo Marques. 2009. A Music Classification Method Based On Timbral Features", International Society for Music Information Retrieval Conference (ISMIR).

[6] P. Dhanalakshmi, S. Palanivel. and V. Ramalingam. 2008. Classification of audio signals using SVM and RBFNN. Expert Systems with Applications Elsevier.

[7] Riccardo Miotto. and Gert Lanckriet. 2012. A Generative Context Model for Semantic Music Annotation and Retrieval. IEEE Transactions on Audio, Speech, And Language Processing. Vol. 20, No. 4, May.

[8] Hung-Yi Lo, Ju-Chiang Wang. and Hsin-Min Wang. 2008. Homogeneous Segmentation and Classifier Ensemble for Audio Tag Annotation and Retrieval. National Science Council of Taiwan.

[9] Guodong Guo. and Stan Z. Li. 2003. Content-Based Audio Classification. and Retrieval by Support Vector Machines. IEEE Transactions on Neural Networks. Vol. 14, No. 1, January.

[10] Dongge Li. and Ishwar K. Sethi. 2001. Classification of General Audio Data for Content Based Retrieval. Pattern Recognition Letter, Elsevier.

[11] Jean-Julien Aucouturier, François Pachet. and Mark Sandler. 2005. The Way It Sounds. Timbre Models For Analysis And Retrieval of Music Signals, IEEE Transactions On Multimedia. Vol. 7, No. 6, December.

[12] George Tzanetakis. 2002. Musical Genre Classification of Audio Signals. IEEE Transactions on Speech And Audio Processing. Vol. 10, No. 5, July.

[13] Santosh V. Chapaneri. 2012. Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping. International Journal of Computer Applications (0975 – 8887). Volume 40– No.3, February.

[14] Tao Li, Mitsunori Ogihara. and Qi Li. 2003. A Comparative Study on Content-Based Music Genre Classification. SIGIR'03.

[15] Chandika Mohan Babu, Manish Puri. and Anamika Das. 2012. Effective principle analysis of speech recognition systems using MFCC and time domain approach for isolated word for training phase spectrum. World Journal of Science and Technology. April.

[16] Atanas Ouzounov. 2010. Cepstral Features and Text-Dependent Speaker Identification –A Comparative Study. Cybernetics and Information Technologies. Volume 10, No 1.

[17] Hyoung-Gook Kim, Nicolas Moreau. and Thomas Sikora. 2004. Audio Classification Based on MPEG-7 Spectral Basis Representations. IEEE Transactions on Circuits and Systems for Video Technology. Vol. 14, No. 5, May.

[18] Emiru Tsunoo, George Tzanetakis, Nobutaka Ono. and Shigeki Sagayama. 2011. Beyond timbral statistics: improving music Classification using percussive Patterns and bass lines. IEEE Transactions on Audio, Speech, And Language Processing. Vol. 19,No. 4, May.