International Journal of Advanced Computer Science and Technology.
ISSN 2249-3123 Volume 6, Number 1 (2016), pp. 13-17

http://www.ripublication.com

Speech/Music Classification using Power Normalized Cepstral Coefficients and K-means

R. Thiruvengatanadhan

Department of Computer Science and Engineering,
Annamalai University, Annamalainagar, Tamil Nadu, India.

Abstract

The objective of a speech/music classification is to classify speech and music data using one or more acoustic characteristics associated with the signal. A speech/music classification system is developed which utilizes the Power Normalized Cepstral Coefficients (PNCC) as the acoustic feature. Multi resolution analysis is the most significant statistical way to extract the features from the input signal and in this study, a method is deployed to model the extracted wavelet feature. k-means clustering organizes the feature vectors into k number of groups. Classifying is done by minimizing the Euclidean distance between feature vector and corresponding cluster centroid.

Keywords: Speech, Music, Feature Extraction, MFCC, K-means.

I. INTRODUCTION

Digital processing is now getting preference for handling audio and speech. This is because new audio applications and systems are predominantly a digital in nature. Speech, Audio and Hearing related research or development are centered on digitalization these days. Since digitalization fosters platform independence, one can create and prototype using a digital processing platform, and then deploy on another platform [1]. Such a development platform would be for ease-of-use and testing, while the criteria for a deployment platform may be totally separate: low power, small size, high speed, low cost, etc.

Digitalized information has shown the problem of automatic audio indexing and classification as essential one for broadcasting process or analysis of stored multimedia data. Numerous researches are investigating these problems nowadays.

The speech/music change point detection is to find the change points between two successive categories (speech and music) present in the audio. This topic has been of a greater interest because it helps in detecting audio category changes, which is an

important pre-processing step. For performing classification, segmented audio is taken as input rather than raw data [6].

Firstly, the audio signal should characterize the audio signal as either one of speech, music or silence [2]. This step can employ any approach like, metric-based, model-based, decoder-guided, model-selection-based and hybrid approaches. Metric-based methods simply measure the difference between two consecutive audio clips that are shifted along the audio signal and speech/music changes are identified at the maxima of the dissimilarity in terms of some distance metric. Decoder guided approach segments a speech stream into male and female clips via a gender-dependent phone recognizer. In model-selection based methods, the segmentation problem is switched to a model selection problem between two nested competing models. Recently, hybrid methods have been given much effort because it combines all the merits from different approaches for giving a better performance.

II. ACOUSTIC FEATURE EXTRACTION

Acoustic feature extraction plays an important role in constructing an audio classification system. The aim is to select features which have large between class and small within class discriminative power.

A. Power Normalised Cepstral Coefficients (PNCC)

PNCC is well known for the high accuracy of automatic speech recognition systems even in high-noise environments [7]. PNCC is an acoustic feature which performs the computation using online algorithms in real-time and provides high accuracy even in noisy conditions [10]. In Fig. 1 Shows the block diagram for the extraction of PNCC features.

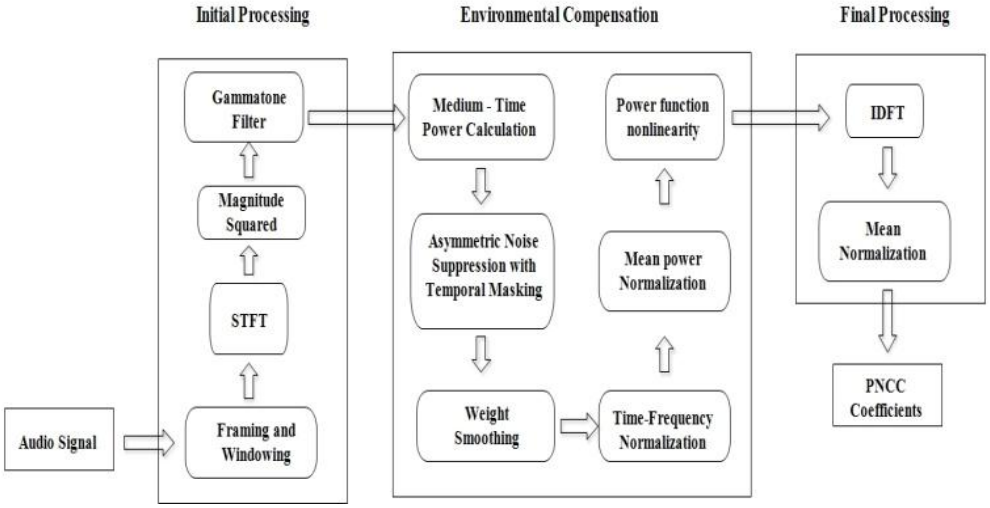

Fig. 1 PNCC Feature Extractions.

Power Normalized Cepstral Coefficients (PNCC) is well known for the accuracy of automatic speech recognition systems, even in high-noise environments [3].

III. TECHNIQUES

A. k-means

Clustering is an unsupervised learning problem which deals with finding a structure in a collection of unlabeled data [4]. It is the process of organizing objects into groups whose members are similar in some way. A cluster is therefore a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters.

K-means algorithm is one of the clustering algorithms that groups data with similar characteristics or features together. These groups of data are called clusters [5]. The data in a cluster will have similar features or characteristics which will be dissimilar from the data in other clusters. K-means clustering organizes the feature vectors into k number of groups. Grouping is done by minimizing the Euclidean distance between feature vector and corresponding cluster centroid. The K-means clustering algorithm is described below:

1. Initialize k centroids.
2. Compute the distance between each feature vector and the centroids.
3. Assign the feature vector to the centroid whose distance is minimum.
4. Re-estimate the centroids.
5. Repeat the above three steps until there is no change in centroids or for a fixed number of iterations.

IV. EXPERIMENTAL RESULTS

A. The database

The speech and music audio data are recorded various sources namely 300 clips of speech and 300 clips of music. Each clip consists of audio data ranging from one second to about ten seconds, with a sampling rate of 8 kHz, 16-bits per sample, monophonic, and 128 kbps audio bit rate.

B. Acoustic feature extraction

13 set of PNCC feature is extracted from each frame of the audio by using the feature extraction techniques. PNCC feature will be calculated for the given wav file. The above process is continued for 600 wav files. The feature values for all the wav files will be stored separately for speech and music.

Experiments were conducted to test the performance of the system using K-means. In this work, K-means clustered gave better performance. Fig. 2 shows the performance of speech and music classification using K-means for different duration respectively.

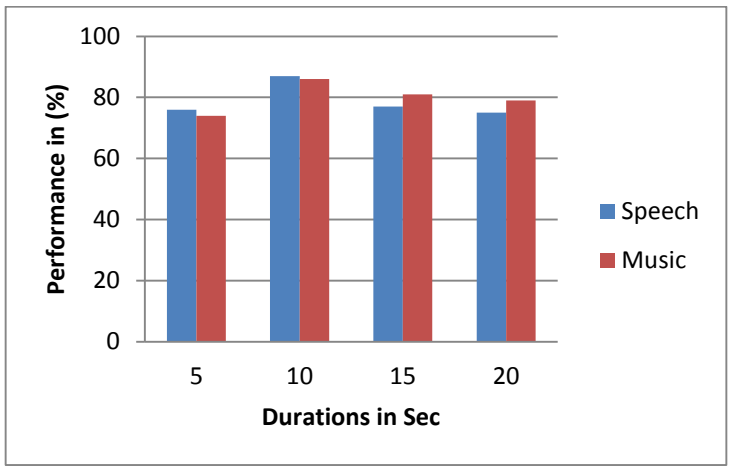

Fig. 2: Performance of audio classification for different duration of speech and music clips using K-means

V. CONCLUSIONS

In this paper, PNCC features for the classification of speech and music files are presented. Further it is possible to improve the classification accuracy by using different types of domain based features together. The proposed classification method is implemented using K-means clustering for classification. The overall accuracy of proposed method K-means using PNCC is 87%. It shows that the proposed method can achieve better classification accuracy.

REFERENCES

[1] Ian Mc Loughlin, Applied Speech and Audio Processing: With MATLAB Examples, Cambridge University Press, 2009.

[2] Chien-Chang Lin, Shi-Huang Chen, Trieu-Kien Truong, and Yukon Chang, "Audio Classification and Categorization Based on Wavelets and Support Vector Machine," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 5, pp. 644-651, September 2005.

[3] Chanwookim and Stern, R. M., "Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4101-4104, 2012.

[4] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," Pattern Recognition, vol. 36, no. 2, pp. 451–461, March 2003.

[5] Konstantin Biatov, "Audio clips retrieval using anchor reference space and latent semantic analysis," in Proc. 11th IEEE Int. Symposium on Multimedia, California,USA, December 2009, pp. 32–37.

[6] K R. M. Aarts and R. T. Dekkers, "A real-timespeech-music discriminator," J. Audio Engi-neering Society, vol. 47, no. 9, pp. 720–725,September 1999.

[7] Xin Yan and Ying Li, "Anti-noise Power Normalized Cepstral Coefficients for Robust Environmental Sounds Recognition in Real Noisy Conditions," Fourth International Conference on Computational Intelligence and Communication Networks, pp. 263-267, 2012.

[8] Chanwoo kim, Stern, R.M. "Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition" IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp:4101 – 4104, 25-30 March 2012.