International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 10 (2017), pp. 2449-2452
© Research India Publications
http://www.ripublication.com

Speech/Music Classification using MFCC and KNN

R. Thiruvengatanadhan

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

Abstract

Today, digital audio applications are part of our everyday lives. Automatic audio classification is very useful in audio indexing; content based audio retrieval and online audio distribution. The accuracy of the classification relies on the strength of the features and classification scheme. In this paper, Mel-Frequency cepstral coefficients (MFCC) features are extracted from the input signal. K-Nearest Neighbour (KNN) is a supervised learning technique where a new instance is classified based on the closest training samples present in the feature space. The proposed KNN model classifies the given input signal is either speech or music.

Keywords: Speech, Music, Feature Extraction, MFCC, KNN.

I. INTRODUCTION

Multimedia databases or file systems can easily have thousands of audio recordings. However, the audio is usually treated as an opaque collection of bytes with only the most primitive fields attached; namely, file format, name, sampling rate, etc. Meaningful information can be extracted from digital audio waveforms in order to compare and classify the data efficiently. When such information is extracted, it can be stored as content description in a compact way. These compact descriptors are of great use not only in audio storage and retrieval applications, but also in efficient content-based segmentation, classification, recognition, indexing and browsing of data. The need to automatically classify, to which class an audio sound belongs, makes audio classification and categorization an emerging and important research area [1]. During the recent years, there have been many studies on automatic audio classification using several features and techniques. A data descriptor is often called a feature vector and the process for extracting such feature vectors from audio is called audio feature extraction. Usually a variety of more or less complex descriptions can be extracted to feature one piece of audio data [2]. Audio refers to speech, music as well as any sound signal and their combination.

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[7].

A. Mel-Frequency cepstral coefficients(MFCC)

MFCCs are short-term spectral features and are widely used in the area of audio and speech processing [3]. The mel frequency cepstrum has proven to be highly effective in recognizing the structure of music signals and in modeling the subjective pitch and frequency content of audio signals. Fig. 1 describes the procedure for extracting the MFCC features.

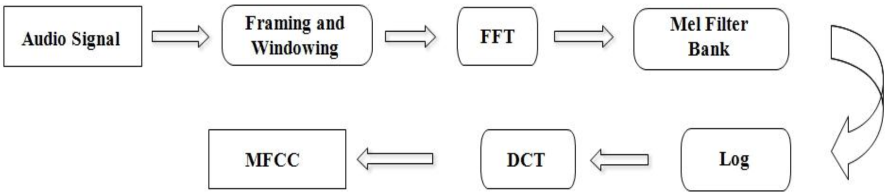

Fig. 1 Extraction of MFCC from Audio Signal.

The MFCCs have been applied in a range of audio mining tasks, and have shown good performance compared to other features. MFCC is computed by various authors in different methods [4]. Computes the cepstral coefficients along with delta cepstral energy and power spectrum deviation while results in 26 dimensional features. The low order MFCCs contains information of the slowly changing spectral envelope while the higher order MFCCs explains the fast variations of the envelope. MFCCs are based on the known variation of the human ears critical bandwidths with frequency, filters spaced linearly at low frequencies and logarithmically at high frequencies to capture the phonetically important characteristics of speech and audio. To obtain MFCCs, the audio signals are segmented and windowed into short frames of 20 msec.

III. TECHNIQUES

A. K-Nearest Neighbour (KNN)

KNN is a supervised learning technique where a new instance is classified based on the closest training samples present in the feature space [5]. It does not use any model to fit, and is only based on memory. When a test data is entered, it is assigned to the class that is most common amongst its k nearest neighbours. KNN classifier is non-parametric method used for classification. It does not need any prior knowledge about, the structure of the data in training set. If the new training pattern is added to existing training set. Any ties can be broken at random. The K-NN algorithm uses the neighborhood classification as the prediction value of the new query instance. The KNN algorithm is sensitive to the local structure of the data. The K-Nearest Neighbor is one of those algorithms that are very simple to understand but works incredibly well in practice [6].

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. The waveform audio format is converted into raw values i.e. 8000 sample values per second.

B. Acoustic feature extraction

13 MFCC features are extracted as a frame size of 20 ms and a frame shift of 10ms of 100 frames as window are used. Hence, an audio signal of 1 second duration results in 100 × 13 feature vector. KNN model is used to classify the acoustic feature vectors. A MFCC feature shows a better classification performance.

Experiments were conducted to test the performance of the system using KNN. In this work, KNN modeled gave better performance. Fig. 2 shows the performance of audio classification using KNN for different duration respectively.

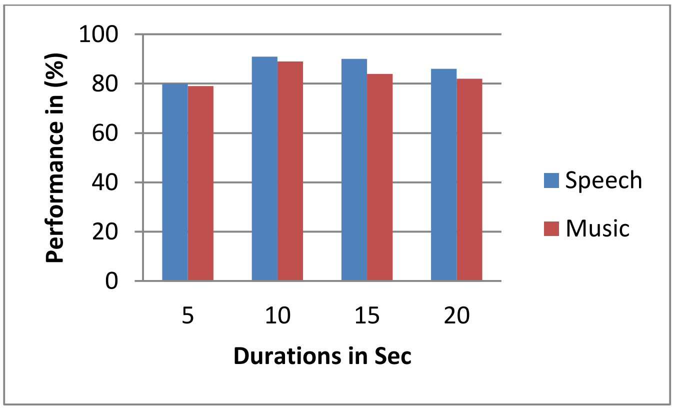

Fig. 2: Performance of audio classification for different duration of speech and music clips using KNN

V. CONCLUSIONS

In this paper, MFCC 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 KNN for classification. The overall accuracy of proposed method

KNN using MFCC is 91%. It shows that the proposed method can achieve better classification accuracy than other approaches. As the classification accuracy is high, this method can retrieve a data more effectively from a large database.

REFERENCES

[1] Toru Taniguchi, MikioTohyama, and Katsuhiko Shirai. Detection of speech and music based on spectral tracking. Speech Communication, 50:547–563, April 2008.

[2] C. Panagiotakis and G. Tziritas. A speech/music discriminator based on rms and zero-crossings,.IEEE Trans. Multimedia, 7(5):155–156, February 2005.

[3] O.M. Mubarak, E. Ambikairajah, and J. Epps, "Novel features for effective speech and music discrimination," in-Proc. 1st IEEE Engineering on Intelligent Systems, Islamabad, Pakistan, April 2006, pp. 342-346

[4] Ahmad R. Abu-El-Quran, Rafik A. Goubran, and Adrian D. C. Chan, "Security monitoring using microphone arrays and audio classification," IEEE Trans. Instrumentation and Measurement, vol. 55, no. 4, pp. 1025-1032, August 2006.

[5] U. Rajendra Acharya, Filippo Molinari, S. Vinitha Sree, Subhagata Chattopadhyay,Kwan-Hoong Nge, and Jasjit S. Suri, "Automated diagnosis of epileptic EEG using entropies," Elsevier Biomedical Signal Processing and Control, vol. 7, pp. 401–408,2012.

[6] Subhas Halder, An Approach to Diagnosis of Cancer Using k-Nearest Neighbor (k-NN)Algorithm, Ph. D thesis, Jadavpur University Kolkata, May, 2013.

[7] C. Panagiotakis and G. Tziritas. A speech/music discriminator based on rms and zero-crossings,.IEEE Trans. Multimedia, 7(5):155–156, February 2005.