Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 11 (2017) pp. 3205-3209

http://www.ripublication.com

Speech/Music Classification using Discrete Wavelet Transform and Linear Discriminate Analysis

R. Thiruvengatanadhan

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

Abstract

This paper presents a new technique to identify the classification of audio data between speech and music category. A speech/music classification system is developed which utilizes the Discrete Wavelet Transform (DWT) 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. Linear Discriminate Analysis (LDA) Linear Discriminate Analysis is one of the most popular and efficient classifier. The proposed LDA model classifies the given input signal is either speech or music.

Keywords: Speech, Music, Feature Extraction, Discrete Wavelet Transform, Linear Discriminate Analysis.

I. INTRODUCTION

Audio refers to speech, music as well as any sound signal and their combination. Audio consists of the fields namely file name, file format, sampling rate, etc. To compare and to classify the audio data effectively, meaningful information is extracted from audio signals which can be stored in a compact way as content descriptors. Meaningful information can be extracted from digital audio waveforms in order to compare and classify the data efficiently [1]. 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.

Music is an art form whose medium is sound and silence. Pitch, rhythm, dynamics, and the sonic qualities are common elements of timbre and texture [2]. The term audio is used to indicate various kinds of audio signals, such as speech, music as well as more general sound signals combinations 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.

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. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is a time scale representation of the audio signal which is computed using digital filtering techniques. DWT is based on subband coding and it reduces the computational time required to yield the wavelet transform. In the early 1970's new techniques were introduced to decompose discrete time signals. Following this, enormous work was done in 1980's in the area of speech signal coding using subband coding and pyramidal coding. After many improvisations the coding schemes led to multi-resolution analysis. DWT feature extraction makes use of filters with varying cut-off frequencies at different scales. An analysing wavelet called mother wavelet implements a prototype function in its wavelet analysis process [3].

The original signal can be developed using a linear combination of wavelet function and appropriate data operations with the wavelet coefficients. The choice of the number of wavelet coefficients depends upon the application and the remaining coefficient below a certain threshold must be truncated. The wavelet scale is computed using subsampling operations namely up and down sampling [4]. The detailed information in the signal is known as the resolution of the signal and is computed using filtering operations.

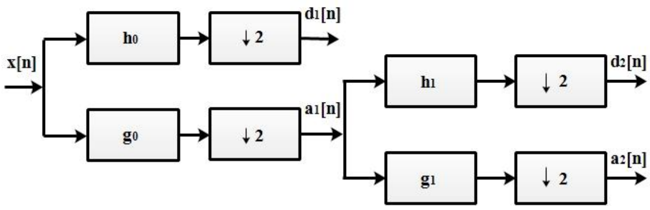

Fig. 1 Two Level Wavelet Decomposition Techniques.

Three mother wavelets namely Haar, Symlets2 and Daubechies8 are best suited for audio processing applications. One approximation and one detailed coefficients are obtained for each frame. By applying DWT, three sets of mother wavelet features are extracted for each frame resulting in a 6 dimensional feature vector.

III. TECHNIQUES

A. Linear Discriminative Analysis(LDA)

LDA classifies a dataset based on the relation between the dispersions within classes and between classes, in order to find the dimension that best classifies a dataset in a linear way [5]. The number of training samples in class, the number of distinct classes, the mean vector of samples belonging to class and represents the set of samples belonging to class. The goal of the LDA is to obtain the matrix that maximizes the relation between classes. The LDA gives the best dimension that describes a dataset dispersion given its features and therefore, is able to analyze the data by reducing its dimensionality [6]. LDA classifier was selected to be used due to its high performance in classification, together with its robustness in long-term use and its low computational cost [7].

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

Six set of features and DWT feature is extracted from each frame of the audio by using the feature extraction techniques. DWT using multi rate filter banks 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.

A DWT feature shows a better classification performance. Daubechies8 wavelet shows high performance when compared to other wavelet as shown in Table 1.

TABLE 1: PERFORMANCE OF SPEECH/MUSIC CLASSIFICATION FOR DIFFERENT WAVELET TRANSFORMS.

Mother wavelet	**Performance(in %)**		
	Speech	**Music**	**Overall**
Haar	88.4	87.4	88.7
Symlets2	86.7	85.6	87.0
Daubechies8	88.7	88.0	89.4

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

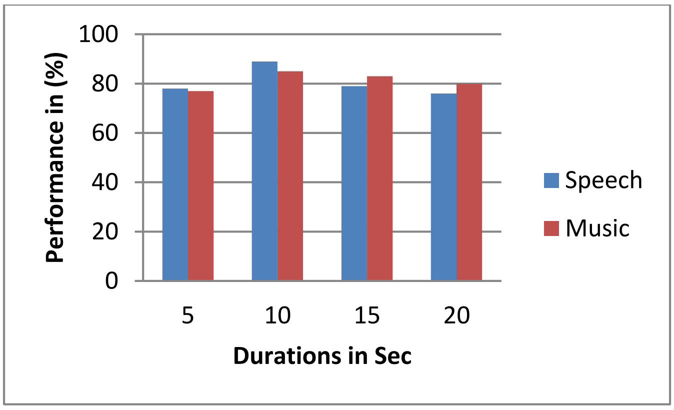

Fig. 5: Performance of audio classification for different duration of speech and music clips using LDA

V. CONCLUSIONS

In this paper, Discrete Wavelet Transform 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 LDA for classification. The overall accuracy of proposed method LDA using DWT is 89%. 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] Chungsoo Lim Mokpo YWL, Chang JH,"Newtechniques for improving the practicality of an svm-based speech/music classifier.",Acoustics, Speech and Signal Processing (ICASSP), pp 1657-1660. 2012.

[2] Lim C, J-H, "Enhancing support vector machine-based speech/music classification using conditional maximum a posteriori criterion.", Signal Processing, IET ,vol. 64, pp 335-340. 2012.

[3] Rekik, S., D. Guerchi, H. Hamam and S.A. Selouani, "Audio Steganography Coding using the Discrete Wavelet Transforms," International Journal of Computer Science Security, vol. 6, pp. 79-83, 2012.

[4] Patil, V. D. and S. D. Ruikar, "Wavelet-Based Image Enhancement using Nonlinear Anisotropic Diffusion," International Journal of Advance Research Computer Science Software Engineering, vol. 2, pp. 158-162, 2012.

[5] Iswarya, P. and Radha, V, "Speech and Text Query Based Tamil -English Cross

Language Information Retrieval system," International Conference on Computer Communication and Informatics, pp. 1-4, Coimbatore, 2014.

[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] K. Englehart, B. Hudgin, and P. A. Parker, "A wavelet-based continuous classification scheme for multifunction myoelectric control," IEEE Transactions on BiomedicalEngineering, vol. 48, no. 3, pp. 302–311, 2001.