Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 01, January-2019

Speech/Music Classification using Subband Coding and AANN

R. Thiruvengatanadhan
Department of Computer Science and Engineering
Annamalai University
Annamalainagar, Tamilnadu, India

Abstract— **The audio refers to speech, music as well as any sound signal and their combination. 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 work, Subband Coding (SBC) features are extracted from the input signal. After feature extraction, classification is carried out, using Autoassociate neural network (AANN) model. The proposed feature extraction and classification models results in better accuracy in speech/music classification.**

Keywords—Feature Extraction, Subband coding, AANN

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. These descriptors are used in segmentation, storage, classification, reorganization, indexing and retrieval 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. The efficiency of a particular feature used for comparison and classification depends greatly on the application, the extraction process and the richness of the description itself. Digital analysis may discriminate whether an audio file contains speech, music or other audio entities. A method is proposed in [2] for speech/music discrimination based on root mean square and zero crossings.

II. SUBBAND CODING

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. Discriminative power of features or feature sets tells how well they can discriminate different classes.

Stress is termed as perceptually induced deviation in the production of speech from that of the conventional production of speech. The excitation plays a vital role in determining the stress information present in the speech signal rather than vocal tract in the linear modeling [3]. Based on the knowledge of stress and its types, the additional information has been incorporated into the speech system which increases the performance of the system.

Subband Coding (SBC) incorporates the excitation in the speech signal whereas mel-scale analysis incorporates properties of human auditory system [4]. In this work a set of features are extracted based on the multi-rate subband analysis or wavelet analysis of stressed speech. The Discrete Cosine Transform (DCT) of subband energy for each frame in the speech signal is extracted using perceptual wavelet packet transform.

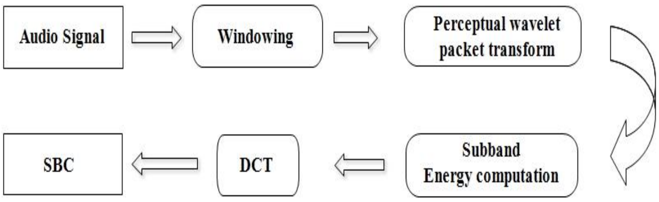
Audio Signal
Windowing
Perceptual wavelet packet transform
Subband Energy computation
DCT
SBC

Fig. 1. SBC Feature Extractions..

This wavelet packet transform can be achieved by two filter banks: low pass filter and high pass filter respectively [5]. The current work is focused to obtain the high energy information in the cascaded filter bank with its wavelet packet tree [6]. Fig.1 shows the block diagram of the extraction procedure of SBC feature.

III. AUTOASSOCIATIVE NEURAL NETWORK (AANN)

Autoassociative Neural Network (AANN) model consists of five layer network which captures the distribution of the feature vector as shown in Fig. 2. The input layer in the network has less number of units than the second and the fourth layers. The first and the fifth layers have more number of units than the third layer [7]. The number of processing units in the second layer can be either linear or non-linear. But the processing units in the first and third layer are non-linear. Back propagation algorithm is used to train the network [8].

The activation functions at the second, third and fourth layer are nonlinear. The structure of the AANN model used in our study is 12L 24N 4N 24N 12L for capturing the distribution of acoustic features, where L denotes a linear unit, and N denotes anon-linear unit. The integer value indicates the number of units used in that layer [9]. The non-linear units use tanh(s) as the activation function, where s is the activation value of the unit. Back propagation learning algorithm is used to adjust the weights of the network to minimize the mean square error for each feature vector [10].

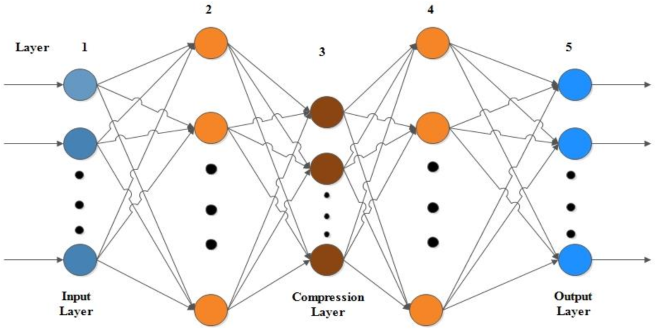

Fig. 2. Auto associate neural network.

IV. EXPERIMENT AND RESULTS

A. The database

Performance of the proposed audio change point detection system is evaluated using the Television broadcast audio data collected from Tamil channels, comprising different durations of audio namely speech and music from 5 seconds to 1 hour. The audio consists of varying durations of the categories, i.e. music followed by speech and speech in between music etc., Audio is sampled at 8 kHz and encoded by 16-bit.

B. Acoustic feature extraction

The feature is extracted from each frame of the audio by using the feature extraction techniques. Here the SBC features are taken. An input wav file is given to the feature extraction techniques. The feature values will be calculated for the given wav file. The feature values for all the wav files will be stored separately for speech and music.

C. Classification

An AANN model is used to capture the distribution of six dimensional spectral and six dimensional of SBC features respectively. The feature vectors are given as input and compared with the output to calculate the error. In this experiment the network is trained for 500 epochs. The confidence score is calculated from the normalized squared error and the category is decided based on highest confidence score. The network structures 12L 24N 4N 24N 12L gives a good performance and this structure is obtained after some trial and error. Fig. 3 shows the performance of AANN for speech/music classification for various durations of training data.

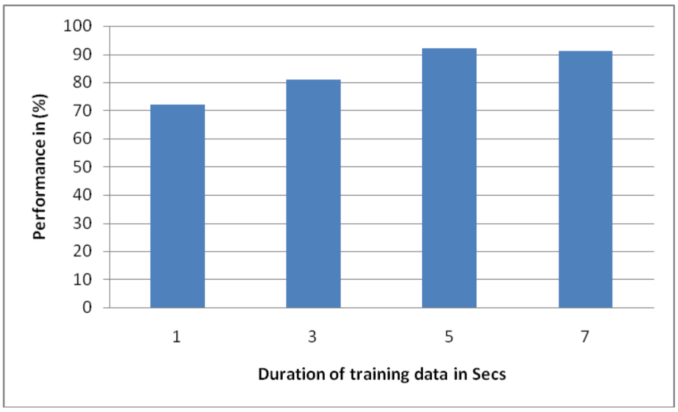

Fig. 3. Performance of AANN for Speech/Music Classification.

V. CONCLUSION

In this paper, we have proposed an speech/music classification system using AANN. SBC is calculated as features to characterize audio content. AANN learning algorithm has been used for the classification of speech and music by learning from training data. The confidence score is calculated from the normalized squared error and the category is decided based on highest confidence score between speech and music by learning from training data. Experimental results show that the proposed audio AANN learning method has good performance in speech/music classification scheme is very effective and the accuracy rate is 92%.

REFERENCES

[1] H Watanabe SM, Kikuchi H (2010) Interval calculation of em algorithmfor gmm parameter estimation. Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium pp 2686–2689

[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] Venkatramaphani kumar S and K V Krishna Kishore, "An Efficient Multimodal Person Authentication System using Gabor and Subband Coding," *IEEE International Conference Computational Intelligence and Computing Research*, pp. 1-5, 2013.

[4] Zhu Leqing, Zhang Zhen "Insect Sound Recognition Based on SBC and HMM," *International Conference on Intelligent Computation Technology and Automation, IEEE*, pp. 544-548, 2010.

[5] Chaya. S, Ramjan Khatik, Siraj Patha and Banda Nawaz, "Subband Coding of Speech Signal Using Scilab", *IPASJ International Journal of Electronics & Communication (IIJEC)*, vol. 2, Issue 5, 2014.

[6] Mahdi Hatam and Mohammad Ali Masnadi-Shirazi, "Optimum Nonnegative Integer Bit Allocation for Wavelet Based Signal Compression and Coding," *Information Sciences Elsevier*, pp. 332-344, 2015.

[7] ShaojunRen, Fengqi Si, Jianxin Zhou, Zongliang Qiao, Yuanlin Cheng, "A new reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems," Chemometrics and Intelligent Laboratory Systems, Volume 172, 15 January 2018, Pages 118-128N.

[8] Nitananda, M. Haseyama, and H. Kitajima, "Accurate Audio-Segment Classification using Feature Extraction Matrix," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 261-264, 2005.

[9] G. Peeters, "A Large Set of Audio Features for Sound Description," Technical representation, IRCAM, 2004.

[10] K. Lee, "Identifying Cover Songs from Audio using Harmonic Representation," International Symposium on Music Information Retrieval, 2006.