International Journal of Computational Science and Engineering.
ISSN 2249-4251 Volume 6, Number 1 (2016), pp. 1-5

http://www.ripublication.com

Speech/Music Change Point Detection using Subband Coding and SVM

R. Thiruvengatanadhan

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

Abstract

The aim of audio change point detection is to check for differences in the distribution of adjacent windows and the transition of the types of audio. In this paper, Subband Coding (SBC) features are extracted which are used to characterize the audio data. Support Vector Machine (SVM) is used to detect change point of audio. The results achieved in our experiments illustrate the potential of this method in detecting the change point between speech and music changes in audio signals.

Keywords: Speech, Music, Feature Extraction, SBC and SVM.

I. INTRODUCTION

The use of digital audio signal processing includes a sequence of processes, for compressing the digitalized audio signal, productions of audio effects as well as classifying the audio. In today’s world, one can observe a great importance of multimedia content management in audio segmentation and classification. The major problems are experienced in audio and visual data handling. It is occupied by audio classification with important applications in broad fields [1].

Audio is a phenomenon that involves various processing. Human perceives the audio in two dimensions, which comprise amplitude and time. The necessity of today is computers that have the hearing capacity of the things that humans can hear. To achieve this, it is important to convert the audio to numerical form. If enough steps are taken in unit time interval, the following problems will occur [2]. At first, frequencies that are higher won’t get grabbed and the computer will not succeed in distinguishing between the pressures that air exerts.

Quantization is done after allocation of little bits by giving more bits to it [3].

The category change point in broadcast audio data consisting of speech and music categories [4]. The category change point detection is made using Sonogram features extracted from the broadcast audio data and techniques such as AANN are used.

II. ACOUSTIC FEATURE EXTRACTION

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

A. Subband Coding (SBC)

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 [5]. 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 [6]. 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.

This wavelet packet transform can be achieved by two filter banks: low pass filter and high pass filter respectively [7]. The current work is focused to obtain the high energy information in the cascaded filter bank with its wavelet packet tree [8]. Fig. 1 shows the block diagram of the extraction procedure of SBC feature.

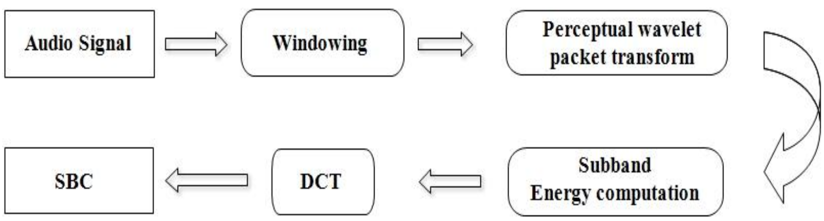

Fig. 1 SBC Feature Extractions.

III. TECHNIQUES

A. Support Vector Machine (SVM)

A machine learning technique which is based on the principle of structure risk minimization is support vector machines. It has numerous applications in the area of pattern recognition [8]. SVM constructs linear model based upon support vectors in order to estimate decision function. If the training data are linearly separable, then SVM finds the optimal hyper plane that separates the data without error.

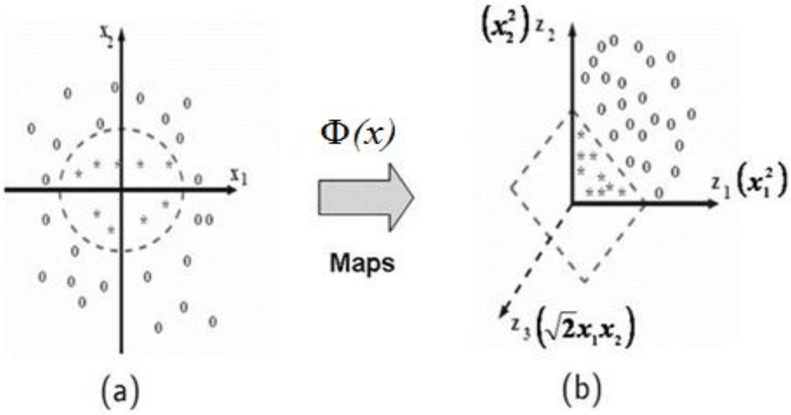

Fig. 2 Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

Fig. 2 shows an example of a non-linear mapping of SVM to construct an optimal hyper plane of separation. SVM maps the input patterns through a non-linear mapping into higher dimension feature space. For linearly separable data, a linear SVM is used to classify the data sets. The patterns lying on the margins which are maximized are the support vectors. The support vectors are the (transformed) training patterns and are equally close to hyperplane of separation. The support vectors are the training samples that define the optimal hyperplane and are the most difficult patterns to classify. Informally speaking, they are the patterns most informative of the classification task [184]. The kernel function generates the inner products to construct machines with different types of non-linear decision surfaces in the input space.

IV. EXPERIMENTAL 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.

B. Acoustic feature extraction

12 SBC features are extracted 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 × 12 feature vector. SVM models are used to capture the distribution of the acoustic feature vectors.

C. Category change point detection

The sliding window is initially placed at the left end of the signal. The classifier SVM model is trained to map the distribution of the feature vectors in the left and right half of the window over the hyper plane, then the misclassification rate of the left and right half feature vectors of the window are used for testing. The category change points are detected from the misclassifications by applying a threshold. A low misclassification indicates that the characteristics of the signal in the right half of the window are different from the signal in the left half of the window, and hence, the middle of the window is a category change point. The performance of the proposed speech/music change point detection system is shown in Fig. 3 for SVM.

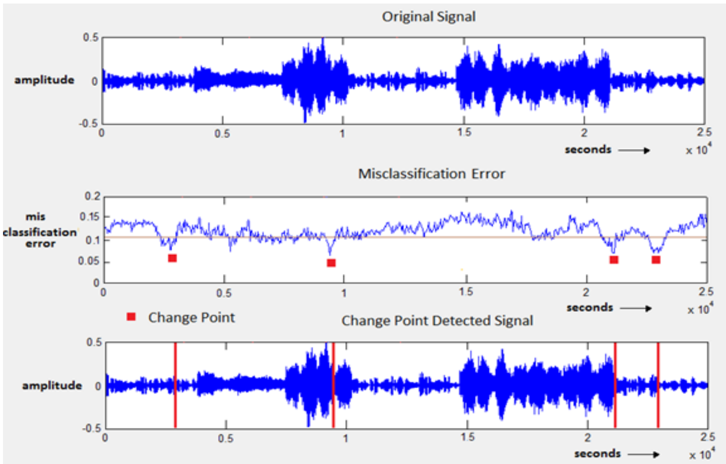

Fig. 3 Snapshot of Speech/Music Change Point Detection Systems Using SVM.

The performance of the speech/music change point detection system using SVM to detect the change point in terms of the various measures is shown in Fig. 4.

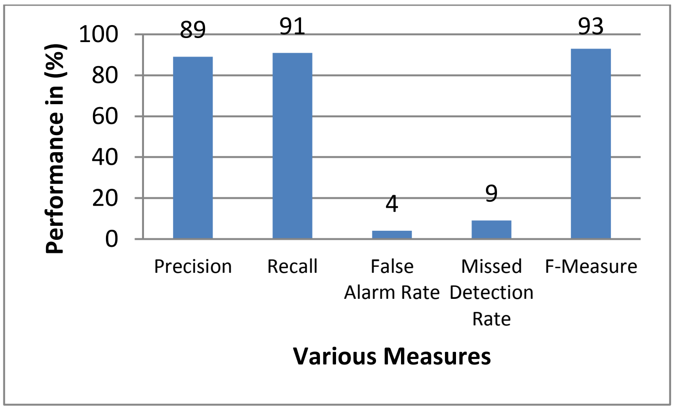

Fig. 4 Performance of to detect the change point in terms of the various measures using SVM.

V. CONCLUSIONS

In this paper we have proposed a method for detecting the category change point between speech/music using SVM. The performance is studied using 12 dimensional SBC features. The performance of the system is evaluated as a large dataset collected from Television broadcast speech/music of various channels. SVM based change point detection gives a better performance of 93% F-measure is achieved.

REFERENCES

[1] J. Aucouturier and F. Pachet, “Representing Musical Genre: A State of Art,” Journal of New Music Research, 2002.

[2] Atti, Andreas Spanias, Ted Painter and Venkatraman, Audio signal processing and coding, Hoboken, NJ: John Wiley & Sons, 2006.

[3] Alnadabi, Muhammad, Saeid and Muhammad, Speech/Music Discrimination: Novel Features in Time Domain, Durham theses, Durham University, 2010.

[4] Francis F. Li, “Nonexclusive Audio Segmentation and Indexing as a Pre-processor for Audio Information Mining,” 26th International Congress on Image and Signal Processing, IEEE, pp: 1593-1597, 2013.

[5] 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.

[6] 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.

[7] 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.

[8] Chungsoo Lim Mokpo, Yeon-Woo Lee, and Joon-Hyuk Chang, “New Techniques for Improving the practicality of a SVM-Based Speech/Music Classifier,” IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1657-1660, 2012.

[9] Theodoros Theodorou, Iosif Mporas and Nikos Fakotakis, “An Overview of Automatic Audio Segmentation,” International Journal of Information Technology and Computer Science, vol.11, pp. 1-9, 2014.