ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Speech/music change point detection using SBC and AANN

R. Thiruvengatanadhan

Assistant Professor, Dept. of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamilnadu, India

ABSTRACT: Category change point detection of acoustic signals into significant regions is an important part of many applications. Changes in audio signal characteristics help in detecting the category change point between different categories. Change point detection has been used extensively in various tasks such as audio classification and audio indexing. In this paper, Subband Coding (SBC) features are extracted which are used to characterize the audio data. Auto associative neural network 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 Recognition, VAD, SBC, AANN

I.INTRODUCTION

A digital audio recording is characterized by two factors namely sampling and quantization. Sampling is defined as the number of samples captured per second to represent the waveform. Sampling is measured in Hertz (Hz) and when the rate of sampling is increased the resolution is also increased and hence, the measurement of the waveform is more precise. Quantization is defined as the number of bits used to represent each sample. Increasing the number of bits for each sample increases the quality of audio recording but the space used for storing the audio files becomes large. Sounds with frequency between 20 Hz to 20,000 Hz are audible by the human ear [1]. Fig. 1 shows the procedure of the proposed method for change point detection of audio signals.

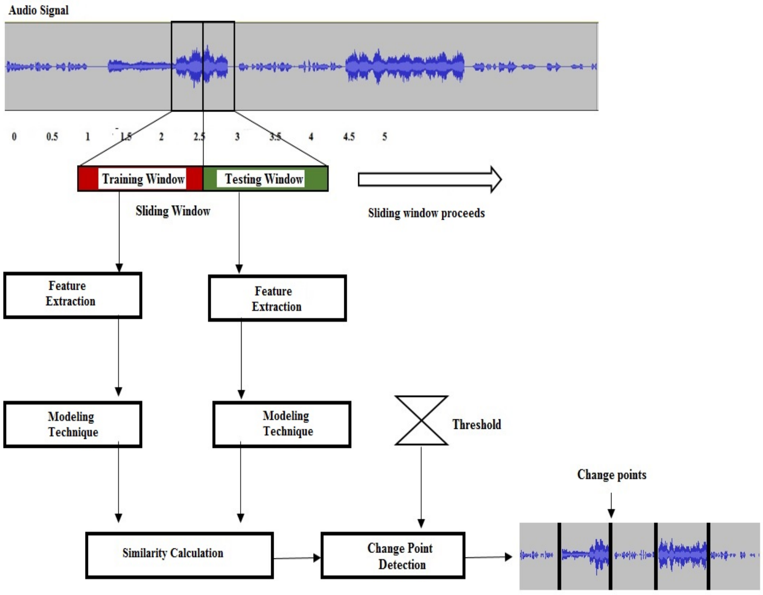

Fig. 1 Proposed Methods for Change Point Detection.

Category change points in an audio signal such as speech to music, music to advertisement and advertisement to news are some examples of segmentation boundaries. Systems which are designed for classification of audio signals into their corresponding categories usually take segmented audios as input. However, this task in practice is a little more complicated as these transitions are not so obvious all the times [2]. For example, the environmental sounds may vary while a news report is broadcast. Thus, many times it is not obvious even to a human listener, whether a category change point should occur or not.

II. 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 [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.

This wavelet packet transform can be achieved by two filter banks: low pass filter and high pass filter respectively [4]. 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.

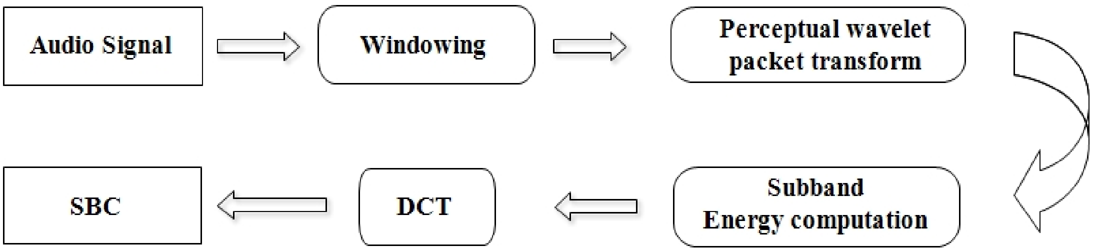
Audio Signal
Windowing
Perceptual wavelet packet transform
Subband Energy computation
DCT
SBC

Fig. 1 SBC Feature Extractions.

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 shape of the hyper surface is determined by projecting the cluster of feature vectors in the input space onto the lower dimensional space simultaneously, as the error between the actual and the desired output gets minimized.

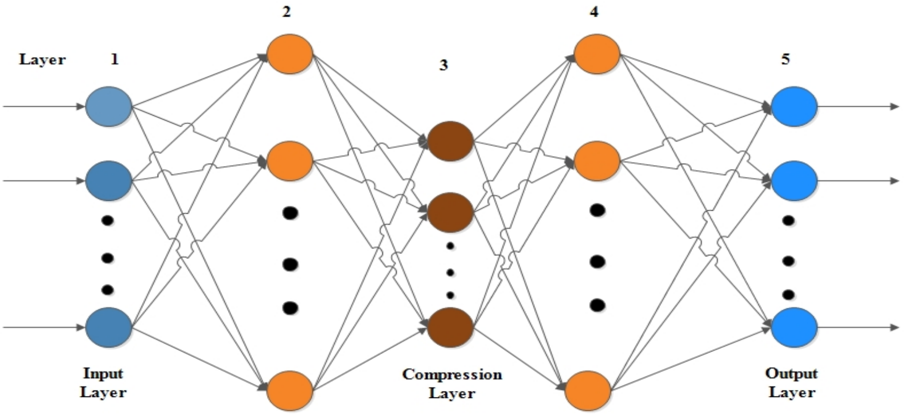

Fig. 2 The Five Layer AANN Model.

During testing the acoustic features extracted are given to the trained model of AANN and the average error is obtained. The structure of the AANN model used in our study is 12L 36N 3N 36N 12L for SBC, for capturing the distribution of the acoustic features.

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. AANN models are used to capture the distribution of the acoustic feature vectors.

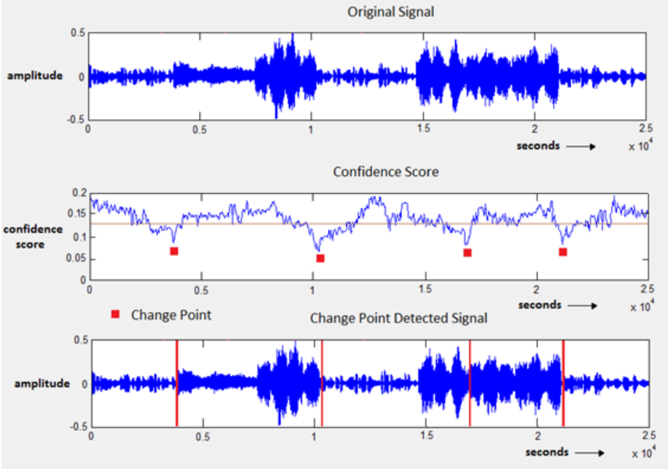

Fig. 3 Snapshot of Speech/Music Change Point Detection Systems Using AANN.

The sliding window of 1 second is initially placed at the left end of the signal. The confidence score for the middle frame of the window is computed by averaging the scores of the frames in the left half of the window. The window is shifted by 10 ms and the same procedure is repeated for the entire signal. The performance of the proposed speech/music change point detection system. Fig. 3 shows a snapshot of Speech/Music Change Point Detection Systems.

The performance of the speech/music change point detection system using AANN to detect the change point in terms of the various measures is shown in Table 1.

Table :1 A comparison of the performance of speech/music Change point detection using AANN in terms of the Various measures.

	Precision	**Recall**	**False Alarm Rate**	**Missed Detection Rate**	**F-Measure**
AANN	90%	84%	9%	14%	86%

VI. CONCLUSION

In this paper we have proposed a method for detecting the category change point between speech/music. The performance is studied using 12 dimensional SBC features using Auto Associative Neural Network (AANN). AANN based change point detection gives a better performance of 86%.

REFERENCES

[1] N. 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.

[2] 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.

[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] D. Li, I. K. Sethi, N. Dimitrova, and T. Mc Gee, “Classification of General Audio Data for Content Based Retrieval,” *Pattern Recognition Letters*, vol. 22, no. 1, pp. 533-544, 2001.

[8] N. 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.