RBFNN and AANN based Speech/music Classification using PNCC

Dr. R. Thiruvengatanadhan
Assistant Professor
Department of Computer Science and Engineering
Annamlai University, Annamalainagar, Tamilnadu, India.
Email Id: thiruvengatanadhan01@gmail.com

Abstract: Audio classification serves as the fundamental step towards the rapid growth in audio data volume. Automatic audio classification is very useful in audio indexing; In this work a speech/music discrimination system is developed which utilizes the Power Normalized Cepstral Coefficients (PNCC) as the acoustic feature. The analysis is done on radial basis function neural network (RBFNN) and Autoassociative Neural Network (AANN) then a conclusion is formed on the basis of their performance and efficiency.

Keywords: Feature Extraction, Pattern Classification, Power Normalized Cepstral Coefficients (PNCC), Radial Basis Function Neural Network (RBFNN), Autoassociative Neural Network (AANN).

I. INTRODUCTION

Acoustics of the fields namely file name, file format, sampling rate, etc. During ongoing years sound characterization is arising as a significant examination region on the grounds that there is a huge need to arrange and to classify the sound information consequently [1]. Audio feature extraction is the process of extracting meaningful information from the audio signal. The features can be more or less complex descriptions and performance of such features depends on the process of extraction [2]. The music signal is a special class in the signal category that has its own characteristics different from the speech signal in many ways. First of all, music normally has a wide range frequency distribution among the audible range of human, from 0 to 20k Hz.

The bandwidth of the speech signal is usually limited into 50 Hz to 7 k Hz and hence, the spectral centroids of music signal are higher than that of the speech. In addition, for considering time-domain characteristics, musical signal usually has a lower silence ratio except that it is sung by a singer or played on a solo instrument only. Compared to an ordinary speech signal, music has lower variability in zero-crossing rate. Besides, music has normally more harmonic than other sound. Therefore, music has higher harmonic than speech. Music usually has regular beats that can be extracted to differentiate it from speech for the sake of the melody and background noise.

II. OUTLINE OF THE WORK

In this study, automatic audio feature extraction and classification approaches are presented. In order to discriminate the speech and music features such as PNCC are extracted to characterize the audio content. RBFNN and AANN are applied to obtain select an optimal models between the classes by learning from training data. Experimental results show that the classification accuracy of RBFNN and AANN with PNCC features can provide a better result. Fig. 1 1illustrates the block diagram of Speech/Music classification system.

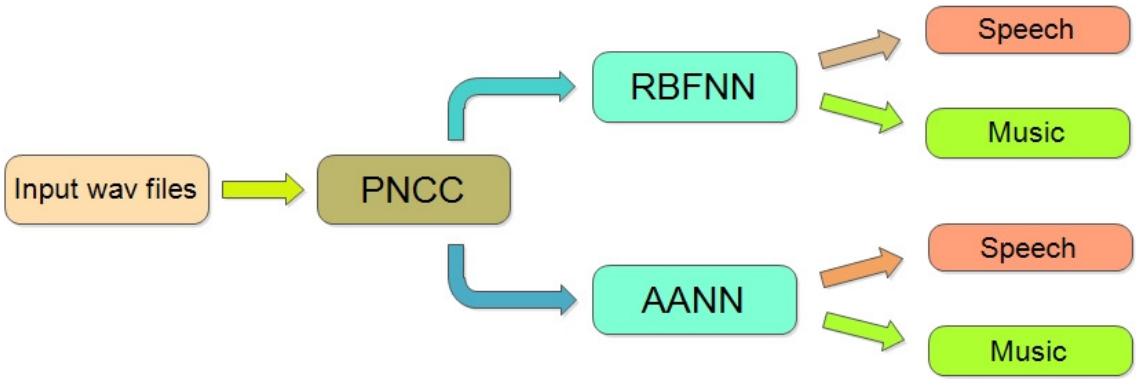

Fig. 1. Block Diagram for speech/music classification.

III. POWER NORMALISED CEPSTRAL COEFFICIENTS (PNCC)

Power Normalised Cepstral Coefficients (PNCC) is well known for the high accuracy of automatic speech recognition systems even in high-noise environments [3]. PNCC is an acoustic element which plays out the calculation utilizing on the web calculations continuously and gives high precision even in loud conditions [4]. It is well known for the accuracy of automatic speech recognition systems, even in high-noise environments. In Fig. 2 Shows the block diagram for the extraction of PNCC features.

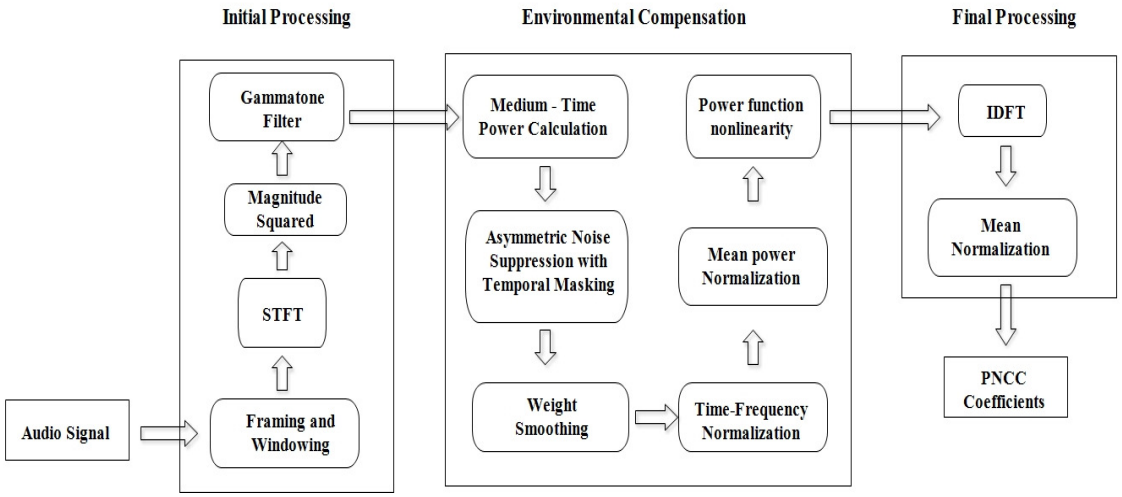

Fig. 2 PNCC Feature Extractions.

IV. RADIAL BASIS FUNCTION NEURAL NETWORK (RBFNN)

It forms a special design with several unique features. A usual RBF neural network classifier has three layers, namely input, hidden, and output layer. Each hidden layer node adopts a radial activated function, and output nodes implement a weighted sum of hidden unit outputs [5]. The output layer is linear, and it produces the predicted class labels based on their sponse of the hidden units. It would be ideal to have them at each distinct point on the input space, but for any realistic problem, only a few input points from all available points are selected using clustering.

V. 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. 3. 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 [6]. During testing the acoustic features extracted are given to the trained model of AANN and the average error is obtained[7]. The structure of the AANN model used in our study is 13L 38N 4N 38N 13L for PNCC, for capturing the distribution of the acoustic features.

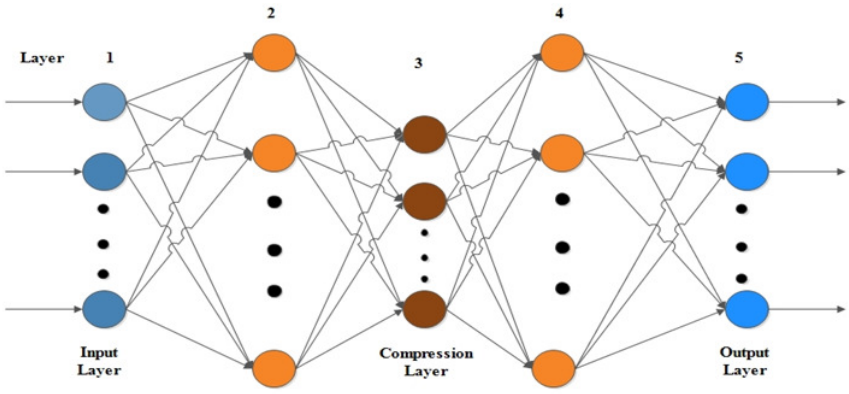

Fig. 3The Five Layer AANN Model.

VI. RESULTS AND DISCUSSION

a. The database

Execution of the proposed sound change point recognition framework is utilizing the Television broadcast sound information gathered from Tamil stations, including various terms of sound specifically discourse and music from 5 seconds to 60 minutes. The sound comprises of fluctuating terms of the classes, for example music followed by discourse and discourse in the middle of music and so on, Audio is tested at 8 kHz and encoded by 16-digit.

b. Acoustic feature extraction

The element is removed from each casing of the sound by utilizing the element extraction strategies. Here the PNCC highlights are taken. An information wav record is given to the component extraction strategies. The component esteems will be determined for the given wav record. The component esteems for all the wav documents will be put away independently for discourse and music.

c. Classification

When the feature extraction process is done the audio should be classified either as speech or music. In a more complex system more classes can be defined, such as silence or speech over musicusing the feature values with appended value RBFNN training is carried out. For testing

the feature extraction is done on different speech and music wav files other than the speech and music wav files used in the training setThe best network was found to be one having 26 basis functions with a learning rate of 0.9 and 0.05 for center and weight respectively. The prediction errors of the validation patterns are larger because these patterns are outside the training space.In Fig. 4 shows Comparison graph for various means in RBFNN.

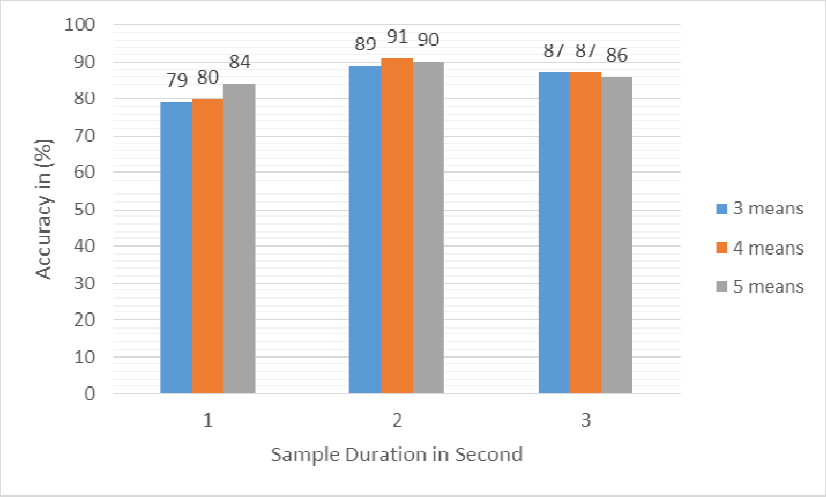

Fig.4 Comparison graph for various means in RBFNN

The network structures 13L 38N 4N 38N 13L gives a good performance and this structure is obtained after some trial and error.In Fig. 5 Performance of AANN for Speech/Music Classification.

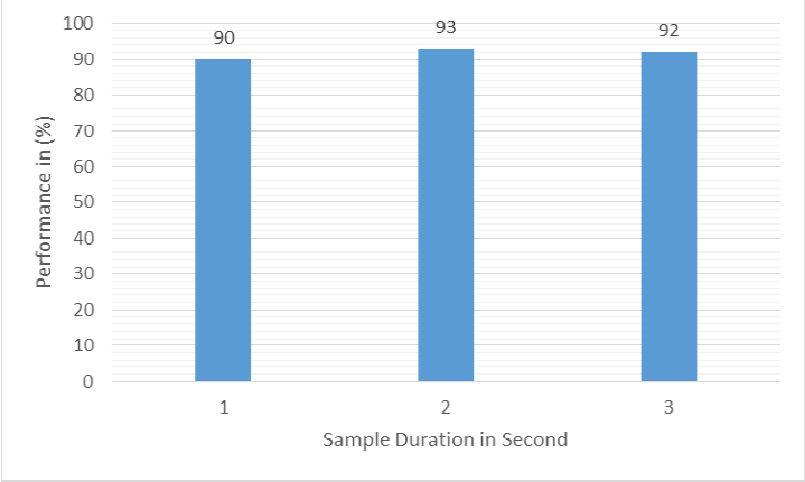

Fig. 5 Performance of AANN for Speech/Music Classification.

VII. CONCLUSION

The system classifies the audio data into speech or music. It is currently the state of the art approach for categorization. In order to classify the audio first the feature extraction is done using PNCC feature. In this paper we have proposed a method for detecting the category change point between speech/music using RBFNN and AANN. The performance is studied

PNCC features. AANN based change point detection gives a better performance of 93% compared with RBFNN.

References

[1] VaishaliJabade, Vedang Deshpande and Aditya K Kumar. Music Generation and Song Popularity Prediction using Artificial Intelligence - An Overview. International Journal of Computer Applications 182(50):33-39, April 2019.

[2] Tayseer M F Taha and Amir Hussain. A Survey on Techniques for Enhancing Speech. International Journal of Computer Applications 179(17):1-14, February 2018.

[3] A. A. Alasadi, R. R. Deshmukh and S. D. Waghmare, "Review of Modgdf& PNCC Techniques for Features Extraction in Speech Recognition," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-7, doi: 10.1109/ICECCT.2019.8869154.

[4] Chanwookim, Stern, R.M. “Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition” IEEE International Conferenceon Acoustics, Speech and Signal Processing (ICASSP), pp:4101 –4104, 25-30 March 2012

[5] D.Tjondronegoro, Y.Chen, and B.Pham, “The power of play break for automatic detection and browsing of self consumable sport video highlights”, In Proceedings of the ACM Workshop on Multimedia Information Retrieval, pp. 267-274, 2004.

[6] 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.

[7] 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