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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 8, August 2017

Speech recognition using PNCC and AANN

R. Thiruvengatanadhan

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

ABSTRACT: Speech recognition is an area of research which deals with the recognition of speech by machine in several conditions. This paper describes a technique that uses Autoassociative Neural Network (AANN) to recognized speech based on features using Power Normalized Cepstral Coefficients (PNCC). Modeling techniques such as AANN were used to model each individual word which is trained to the system. Each isolated word Segment using Voice Activity Detection (VAD) from the test sentence is matched against these models for finding the semantic representation of the test input speech. Experimental results of AANN shows good performance in recognized rate.

KEYWORDS: Speech Recognition, VAD, PNCC, AANN

I.INTRODUCTION

Research in Speech Recognition by machines had been done for almost four decades. Over the past decades, the development of speech recognition applications gives invaluable contributions to this field of research and is becoming more mature in recent years. Various research and development has been done in recent years in speech recognition in various languages [1]. Human speech perception is bimodal in nature which humans combine audio and visual information in deciding what has been spoken, especially in noisy environments [2]. Furthermore, bimodal fusion of audio and visual information in perceiving speech has been shown to be useful for improving the accuracy of speech recognition in both humans and machines by Christian Benoit [3]. In addition, visual speech is of particular importance modality to the hearing impaired person as mouth movement is well known to play an important role in both sign language and simultaneous communication between the deaf [4].

II. VOICE ACTIVITY DETECTION

Voice Activity Detection (VAD) is a technique for finding voiced segments in speech and plays an important role in speech mining applications [5]. VAD ignores the additional signal information around the word under consideration. It can be also viewed as a speaker independent word recognition problem. The basic principle of a VAD algorithm is that it extracts acoustic features from the input signal and then compares these values with thresholds usually extracted from silence. Voice activity is declared if the measured values exceed the threshold. Otherwise, no speech activity is present [6].

VAD finds its usage in a variety of speech communication systems like coding of speech, recognizing speech, hands free telephony, audio conferencing, speech enhancement and cancellation of audio [7], [8]. It identifies where the speech is voiced, unvoiced or sustained and makes smooth progress of the speech process. Fig. 1 shows the isolated word separation.

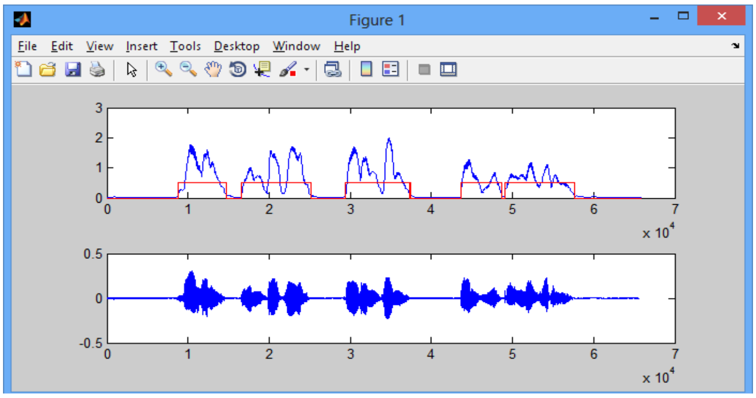

Fig. 1 Isolated Word Separations.

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 [9]. PNCC is an acoustic feature which performs the computation using online algorithms in real-time and provides high accuracy even in noisy conditions [10]. (PNCC) is well known for the accuracy of automatic speech recognition systems, even in high-noise environments. In Fig. 1 Shows the block diagram for the extraction of PNCC features.

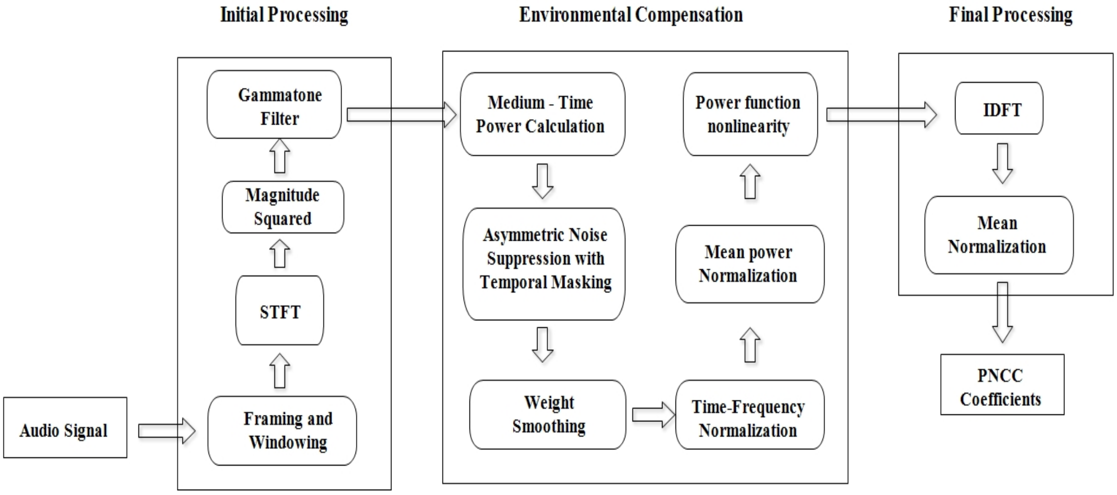

Fig. 1 PNCC Feature Extractions.

IV. 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 [11]. 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 [12].

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.

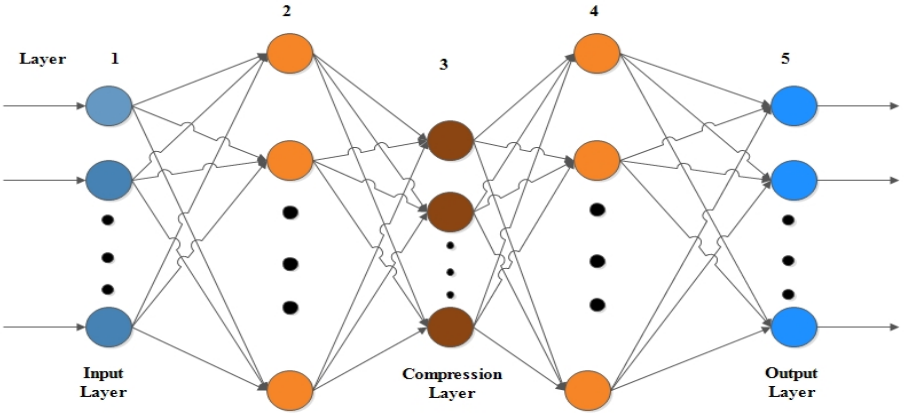

Fig. 2 The Five Layer AANN Model.

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 13L 38N 4N 38N 13L for MFCC, for capturing the distribution of the acoustic features.

V. EXPERIMENTAL RESULTS

A. The database

Experiments are conducted for speech recognition audio using Television broadcast speechdata collected from Tamil news channels using a tuner card. A total dataset of 100 different speech dialogue clips, ranging from 5 to 10 seconds duration, sampled at 16 kHz and encoded by 16-bit is recorded. Voice activity detection is performed to isolate the words in each speech file using RMS energy envelope.

B. Acoustic feature extraction

In this work the pre-emphasized signal containing the continuous speech is taken for testing. Through VAD the isolated words are extracted from the sentences. Thus frames which are unvoiced excitations are removed by thresholding the segment size. Feature PNCC are extracted from each frame of size 320 window with an overlap of 120 samples. Thus it leads to 13 PNCCs respectively which are used individually to represent the isolated word segment. During training process each isolated word is separated into 20ms overlapping windows for extracting 13 PNCCs features.

Using VAD isolated words in a speech is separated. For training, isolated words from were considered. The training process analyzes speech training data to find an optimal way to classify speech frames into their respective classes. 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 performance of speech recognition is studied by varying the number of units in the compression layer as shown in Fig. 3.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 8, August 2017

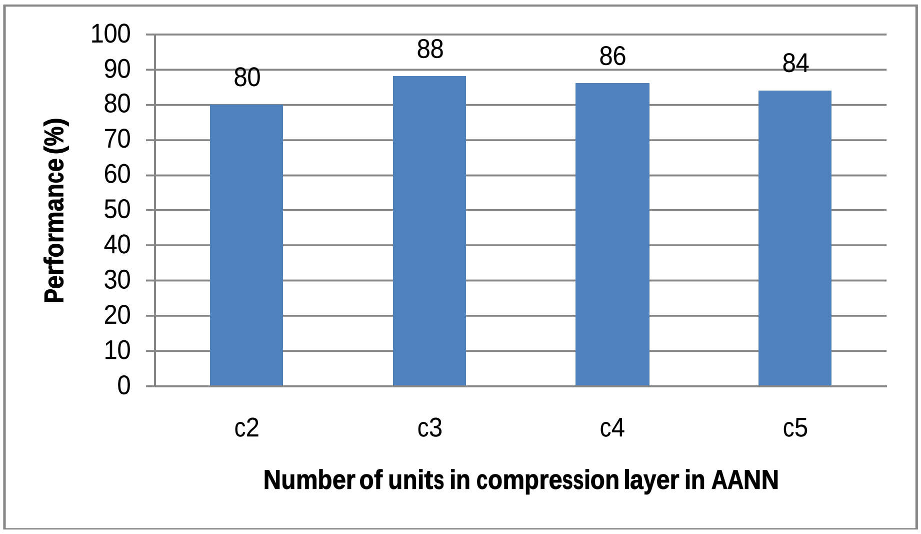

Fig. 3 Performance of Speech Recognition in Terms of Number of Units in the compression Layer

The performance of speech recognition in terms of number of units in the expansion layer is shown in Fig. 4. The network structures 13L 26N 4N 26N 13L gives a good performance and this structure is obtained after some trial and error.

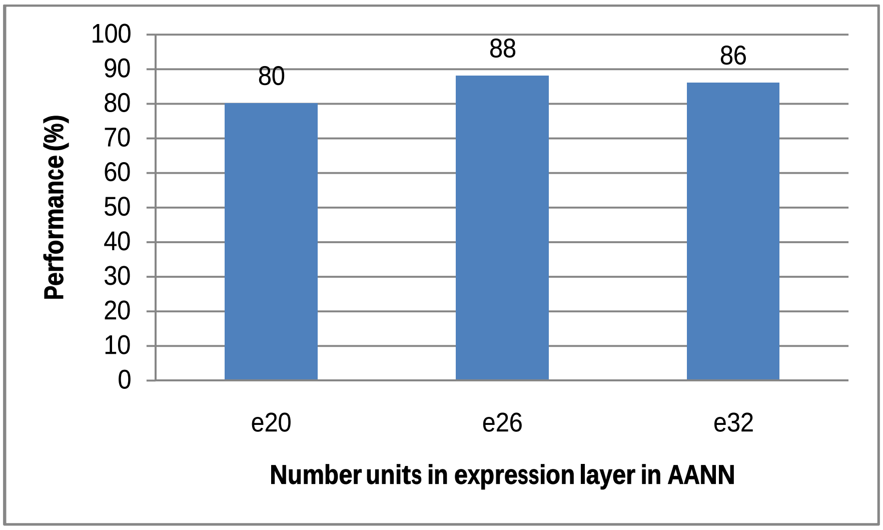

Fig. 4 Performance of Speech Recognition in Terms of Number of Units in the Expansion Layer.

VI. CONCLUSION

In this paper, we have proposed speech recognition system using AANN. Voice Activity Detection (VAD) is used for segregating individual words out of the continuous speeches. Features for each isolated word are extracted and those models were trained successfully. AANN is used to model each Individual utterance. PNCC is calculated as features to characterize audio content. AANN learning algorithm has been used for the recognized speech by learning from

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 8, August 2017

training data. Experimental results show that the proposed audio AANN learning method has good performance in 88% speech recognized rate.

REFERENCES

[1] C. Y. Fook, M. Hariharan, Sazali Yaacob, Adom AH," A Review: Malay Speech Recognition and Audio Visual Speech Recognition" International Conference on Biomedical Engineering (ICoBE),27-28 February 2012,Penang

[2] Gerasimos, P., Chalapathy, N., Juergen, L. and Iain, M., "Audio-Visual Automatic Speech Recognition: An Overview," Issues in Visual and Audio-Visual Speech Processing, MIT Press, 2004.

[3] Benoît, C., "The intrinsic bimodality of speech communication and the synthesis of talking faces". Journal on Communications, 43, Scientific Society for Telecommunications, Hungary (July-Sept), 32-40, 1992.

[4] Marschark, M., LePoutre, D., and Bement, L., "Mouth movement and signed communication," In Campbell, R., Dodd, B., and Burnham, D. (Eds.), Hearing by Eye II. Hove, United Kingdom: Psychology Press Ltd. Publishers, pp. 245–266, 1998.

[5] Ivan Markovi, Sre´ ckoJuri´ Kavelj and Ivan Petrovi, "Partial Mutual Information Based Input Variable Selection for Supervised Learning Approaches to Voice Activity Detection," *Applied Soft Computing Elsevier*, vol. 13, pp. 4383-4391, 2013.

[6] Khoubrouy, S. A. and Panahi, I.M.S., "Voice Activation Detection using Teager-Kaiser Energy Measure," *International Symposium on Image and Signal Processing and Analysis*, pp. 388-392, 2013.

[7] El Bachir Tazi, Abderrahim Benabbou and Mostafa Harti, "Voice Activity Detection for Robust Speaker Identification System," *IJCA Special Issue on Software Engineering, Databases and Expert Systems*, pp. 35-39, 2012.

[8] Nitin N Lokhande, Navnath S Nehe and Pratap S Vikhe, "Voice Activity Detection Algorithm for Speech Recognition Applications," *IJCA Proceedings on International Conference in Computational Intelligence*, vol. 6, pp. 5-7, 2012.

[9] Xin Yan and Ying Li, "Anti-noise Power Normalized Cepstral Coefficients for Robust Environmental Sounds Recognition in Real Noisy Conditions," Fourth International Conference on Computational Intelligence and Communication Networks, pp. 263-267, 2012.

[10] Chanwoo kim, Stern, R.M. "Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition" IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp:4101 –4104, 25-30 March 2012.

[11] 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.

[12] 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.

DOI: 10.15680/IJIRCCE.2017. 0508058