Speaker Identification Systems using AANN with PNCC

T. S. Mullaivendan
Research Scholar
Department of Computer & Information Science
Annamalai University, Annamalainagar
Tamilnadu, India

Dr. R. Thiruvengatanadhan
Assistant Professor
Department of Computer Science & Engineering
Annamalai University, Annamalainagar
Tamilnadu, India

Dr. P. Dhanalakshmi
Professor
Department of Computer Science & Engineering
Annamalai University, Annamalainagar
Tamilnadu, India

Abstract- Speech contains many features or characteristics that can discriminate the identity of the person. In this paper to develop a speaker identification system, which consists of a recorded voice signal, extracting its features and matching it with stored templates. Power Normalized Cepstral Coefficients (PNCC) is applied for feature extraction purpose. Autoassociative Neural Network (AANN) algorithm is used for generating template and feature matching purpose.

Keywords – Speaker identification, Feature Extraction, Power Normalized Cepstral Coefficients (PNCC) and Autoassociative Neural Network (AANN)

I. INTRODUCTION

The speaker identification task results in the best match for an unknown voice from the database, whereas speaker verification task is either to accept/reject a claimed identity. Based on the languages used for the study, speaker recognition can be done in monolingual, crosslingual and multilingual modes [1]. In the monolingual mode, same language is considered for both training and testing. In crosslingual speaker recognition, training and testing are carried out with different languages. In multilingual mode, speaker-dependent models are trained using single language and tested for several languages [2]. Speaker identification refer to the task that is interested in finding identity of the anonymous speakers [3]. Speaker identification is the process of automatically recognizing a speaker by machine using the speaker's voice [4]. The most well-known utilization of speaker recognizable proof frameworks is in access control, for instance, admittance to a room or advantaged data via phone [5]. Additionally it has an exceptionally valuable utilization for speaker transformation in programmed discourse acknowledgment framework.

As speech interaction with the computers become more pervasive in activities such as telephone transactions and information retrieval from speech databases, the ability to automatically recognize a speaker based on his vocal characteristics becomes more useful. The problem of speaker identification can be formulated as a pattern classification problem and methods from statistics and machine learning are suitable. The feature extraction module receives small segments of speech with lengths of audio. From these frames, features that are intended to be unique for a speaker are extracted. Extracted from the frames of audio, features are used to modeling the speaker are shown in the Figure 1.

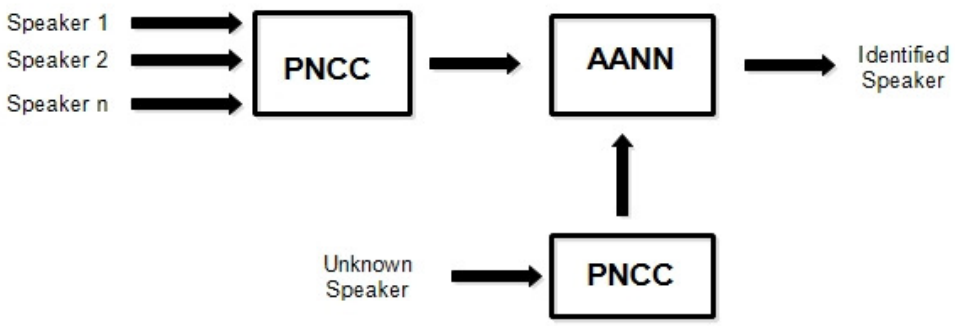

Figure 1. The Speaker Identification System

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 [6]. 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 [7]. 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. It identifies where the speech is voiced, unvoiced or sustained and makes smooth progress of the speech process.

III. POWER NORMALIZED CEPSTRAL COEFFICIENTS (PNCC)

PNCC is one of the most recently developed and most accurate feature which outperforms almost all the other types of conventional features even in highly noisy environments [8]. The higher accuracy achieved by PNCC is mainly attributed to the key features such as the use of a power-law nonlinearity, an asymmetric noise suppression module and temporal masking. The actual human auditory filters, which have non-linearly increasing bandwidths, is represented by a bank of Gammatone filters [9] [10]. PNCC algorithm for extracting features for speech recognition. For this technique, accuracy and complexity was checked which results in increased accuracy because SNR having greater value [11]. Figure 2 shows the PNCC feature extraction.

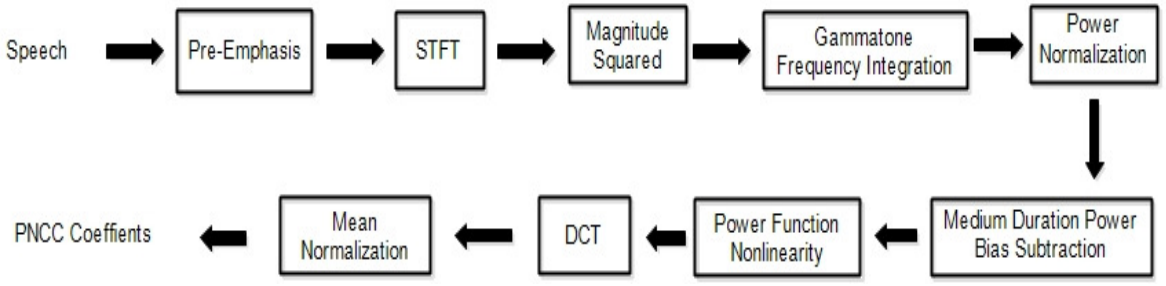

Figure 2. PNCC Feature Extraction

IV. AUTOASSOCIATIVE NEURAL NETWORK MODELS (AANN)

AANN are feed forward neural networks performing an identity mapping of the input space, and are used to capture the distribution of the input data [12],[13]. Let us consider the five layer AANN model, which has three hidden layers. In this network, the second and fourth layers have more units than the input layer. The third layer has fewer units than the first or fifth [14], [15]. The processing units in the first and third hidden layer are nonlinear, and the units in the second compression/hidden layer can be linear or nonlinear. As the error between the actual and the

desired output vectors is minimized, the cluster of points in the input space determines the shape of the hypersurface obtained by the projection onto the lower dimensional space.

V. EXPERIMENT RESULTS

Experiments are conducted for speaker identification, the various speaker voice signal is recorded. Voice samples are was stored in different folder for training and testing. A total dataset of 30 different speaker of 300 voice clips, ranging from 5 to 10 minutes 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. For each speech file, a database of the isolated words is obtained using VAD.

PNCC algorithm for extracting features for speech identification. A frame size of 20 ms with a frame shift of 10 ms is used. Thereby 13 PNCC features are extracted in the speech database.

An AANN model is used to capture the distribution of 13 dimensional of PNCC 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 13L 38N 6N 38N 13L gives a good performance and this structure is obtained after some trial and error. Table 1 shows the performance of the speaker identification system.

The performance of speaker identification system is measured using correct identification rate (CIR), false acceptance rate (FAR) and false rejection rate (FRR) [16].

$$CIR = \frac{Number\ of\ correctly\ identified\ claims}{Total\ number\ of\ claims} \times 100\% \quad (1)$$

$$FAR = \frac{Number\ of\ wrong\ identified\ claims}{Total\ number\ of\ claims} \times 100\% \quad (2)$$

$$FRR = \frac{Number\ of\ wrong\ rejected\ claims}{Total\ number\ of\ claims} \times 100\% \quad (3)$$

Table - 1 The performance of the speaker identification system

Number of Speakers	CIR in %	FAR in %	FRR in %
10	84	5	3
20	89	7	5
30	91	10	9

VI. CONCLUSION

This paper implemented an approach based on AANN for speaker identification. VAD algorithm is that it extracts acoustic features from the input signal and then compares these values with thresholds usually extracted from silence. The PNCC technique has been applied for feature extraction. The focus of this work is mainly on applications which require high identification rates and less computation using approach to take advantage of AANN approaches. The speaker identification accuracy is 91% using AANN.

REFERENCES

[1] P. H. Arjun, Speaker recognition in Indian languages: A feature based approach, Ph.D. dissertation, Indian Institute of Technology Kharagpur, Dept. of Electrical Engg., Kharagpur, India, Jul. 2005.

[2] B. G. Nagaraja, Multilingual speaker identification, Ph.D. dissertation, Visvesvaraya Technological University, Belagavi, India, Oct. 2014.

[3] Feras E. Abualadas, Akram M. Zeki, Muzhir Shaban Al-Ani, Az-Eddine Messikh, "Speaker Identification based on Hybrid Feature Extraction Techniques", International Journal of Advanced Computer Science and Applications,Vol. 10, No. 3, 2019.

[4] Rabiner, L.R., A tutorial on Hidden Markov Models, Proceeding of IEEE, Vol.73, pp.1349-1387, 1989.

[5] U. Bodenhausen, S. Manke, and A. Waibel, Connectionist Architectural Learning for High Performance Character and Speech Recognition, Proceedings of the International Conference on Acoustics Speech and Signal Processing, vol. 1, pp. 625-628, Minneapolis, MN, USA, April 1993.

[6] 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.

[7] 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.

[8] Chanwoo Kim and Richard M. Stern. Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition.IEEE/ACM Transac-tions on Audio, Speech and Language Processing2016;24:1315-1329.

[9] Malcolm Slaney.An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. Apple Computer Technical Report No. 35; 1993.

[10] Badi, Alzahra ; Ko, Kyungdeuk ; Ko, Hanseok. " Bird sounds classification by combining PNCC and robust Mel-log filter bank features." Journal of the Acoustical Society of Korea. 2019 ; Vol. 38, No. 1. pp. 39-46.

[11] P. P. Dani, M. S. Deole, "Speech recognition using PNCC," IEEE ICASSP, vol. 9, pp. 223-231, 2017.

[12] R.Thiruvengatanadhan, P.Dhanalakshmi, S.Palanivel, "Speech/Music Change Point Detection using AANN and GMM", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 9 No. 21, 2014

[13] P. Dhanalakshmi, S. Palanivel, V. Ramalingam," Classification of audio signals using AANN and GMM", Applied Soft Computing, 11 (2011) 716–723

[14] S. Palanivel, Person authentication using speech, face and visual speech, Ph.D. Thesis, IIT, Madras, 2004.

[15] B. Yegnanarayana, S. Kishore, AANN: an alternative to GMM for pattern recog-nition, Neural Networks 15 (2002) 459–469.

[16] S. A. Angadi and S. M. Hatture, "A novel spectral graph theoretic approach to user identification using hand geometry," International Journal of Machine Intelligence, vol. 3, no. 4, pp. 282-288, 2011.