Online Text-Independent Speaker Verification System at IITM

S.P.Kishore and B.Yegnanarayana
Department of Computer Science and Engineering
Indian Institute of Technology, Madras, 600 036, India
E-mail: {kishore@speech., yegna@speech.}iitm.ernet.in

Abstract—**In this paper, we present online text-independent speaker verification system developed at IIT Madras. Unlike, parametric approaches to model the distribution of speaker-specific features, we explore nonparametric approaches to capture the distribution of speaker-specific features. The distribution capturing ability of NonLinear Principal Component Analysis (NLPCA) neural network is exploited to build speaker models. In this study, linear prediction cepstral coefficients are used to represent speaker-specific information. Background normalization is performed using an rank based approach. The effectiveness of this speaker verification system is demonstrated on a database of 15 speakers.**

Keywords— **Speaker verification, Autoassociative neural network models**

I. Introduction

Speaker verification is to verify the speaker from his/her voice. Applications of speaker verification are wide ranging including banking transactions over a telephone network, voice mail, security control for confidential areas and access to the computers. The task of verifying the speaker may be text-dependent or text-independent. For text-dependent cases template matching techniques like dynamic time warping are used, whereas, for text-independent cases, the distribution of speaker-specific features extracted from the speech signal is captured [5] [9]. Most of the current speaker verification systems use Gaussian mixture models to model the distribution of speaker-specific features [1]. Recently, Auto-Associative Neural Network (AANN) models performing NonLinear Principal Component Analysis (NLPCA) are proposed to capture the distribution of speaker-specific features [6] [8]. In this paper, we present the IITM online speaker verification based on AANN models performing NLPCA. The development of this speaker verification system is primarily motivated by our participation in the NIST-99 speaker evaluation [1]. The paper is organized as follows: Section 2 describes the distribution capturing ability of AANN models performing NLPCA. Section 3 describes the IITM online speaker verification. In Section 4, the performance of the online speaker verification system is discussed.

II. Auto-Associative Neural Networks Performing NLPCA

Neural network models performing NLPCA are conventional feedforward networks as shown in Fig.1 [12] [7]. The processing units in the input and output layers are linear, whereas the units in the hidden layer are nonlinear. The network architecture of these models have more than one hidden layer. One of the hidden layers known as bottleneck hidden layer has smaller dimension than either input or output. The network is trained to perform identity mapping using backpropogation learning algorithm. When the network finds an acceptable solution at the output layer, the data compression in the bottleneck hidden layer may force the hidden units to represent the significant features of the data like in Principal Component Analysis (PCA). Since, the processing units in the hidden layers are nonlinear, the projection of the input space R^M onto the lower dimensional space R^N is nonlinear. The hypersurface obtained by projecting the lower dimensional space R^N

back onto the input space R^M will be on the surface of maximum variance of the data. In other words, AANN models can be used to capture the distribution of the given data set [8] [6]. Fig.2(a) shows the artificial 2-D data. An AANN model is trained with the data shown in Fig.2(a). Probability distribution of the data derived from the training error can be seen in Fig.2(b) [3].

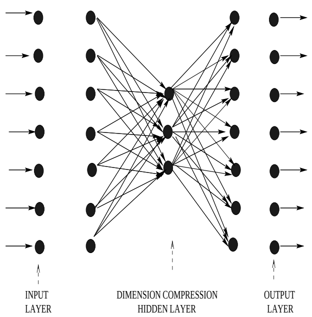
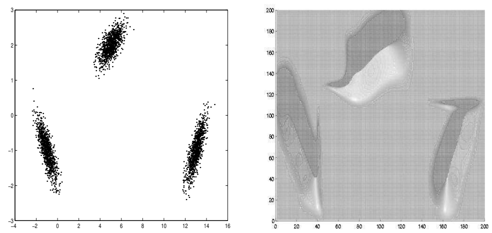


Fig. 1. An Auto-Associative Neural Network



(a) (b)

Fig. 2. (a) Artificial data (b) Distribution captured by the AANN

III. Overview of Online Speaker Verification System

Feature extraction, generation of speaker models and verification procedure are the three major modules of the online speaker verification system. The following sections describe each one of these modules in detail.

A. Feature Extraction

Speech signal of one minute duration is recorded in laboratory environment using the recording facilities of a computer system. The speech signal is sampled at 11025 Hz and the differenced speech is considered for further processing. The feature vectors extracted from the differenced speech signal are 19-dimensional, weighted cepstral coefficients [4]. These cepstral coefficients are obtained from the 16 predictor coefficients using 20ms frames for every 10ms shift. Silence frames are removed using an amplitude threshold. The cepstral coefficients characterize the spectrum of the speech signal which is primarily determined by the vocal cavities and the articulators [5] [2]. The spectral information of the same sound uttered by two persons differ due to the physical differences between the vocal organs and the manner in which they use them during speech production [2]. Since it is difficult to store the templates of all sounds spoken by the speaker, the distribution of the speaker-specific features extracted from the speech signal is captured. In this context, the nonlinear models like AANN are used to capture the complex distribution of the speaker-specific features.

B. Generation of speaker models

The property of AANN models to capture the distribution of given data set is used in building the speaker models. Each speaker model is built by training an AANN with the feature vectors extracted from the utterance of a particular speaker. The structure of the AANN model used is $19L14N32N22N19L$, where L denotes linear units and N denotes nonlinear units. The integer value denotes the number of units in that particular layer. The AANN model is trained in pattern mode using backpropogation learning algorithm for 60 epochs [12]. The total time taken to train the model including feature extraction is approximately 5min (on Pentium III processor).

C. Verification Procedure

During testing, the claimant has to speak for 10sec. The feature vectors extracted from the test utterance are given to the claimant model. The score obtained by the model is the error (Euclidean distance between desired and actual output) averaged over all the feature vectors. The score of the claimant model is compared with a threshold to accept or reject the speaker. Since, the score of the claimant model is sensitive to the linguistic content of the utterance and to the variations of the speaker's voice, normalization of background model is used [10]. Two approaches to represent the background model for AANN-based speaker verification system are compared in [11]. The two approaches are Universal Background Model (UBM) and Individual Background Model (IBM). In IITM online speaker verification system, the approach of IBM is incorporated because of its simplicity in choosing the parameters and effectiveness in performance [11].

The normalization procedure of IBM involves generation of speaker models (other than the registered speakers) called as individual background models. The individual background models are generated with the same parametric specifications used to generate the registered speaker models. In the testing phase, the feature vectors of the test utterance are given to the claimant model and to all the individual background models. The position (R) of the claimant model score in the ascending list of the scores from individual background models is used for accepting/rejecting the speaker.

IV. Results and Discussion

Performance of IITM online speaker verification system is evaluated on 15 cooperative speakers in terms of False Acceptance (FA) and False Rejection (FR). The probability of false acceptance will decrease by using large number of individual background models, but it may increase the probability of false rejection because of genuine models performing poorly. In this evaluation, we have considered 25 models as individual background models.

Table 1 shows the performance of IITM online speaker verification system. From Table 1, we can observe the imposter rejection capability of AANN-based speaker verification system. When training and testing is done using different microphones, the false rejection rate of 15% indicates the effect of microphone on the performance of speaker verification system. Moreover, the limitation of segmental features is observed, when the speaker verification system could not authenticate the genuine claimant, when he/she babbles or partially masks the mouth.

TABLE I
Performance of online speaker verification system

Claimant	No. of tests	Microphone for training & testing	FA in %	FR in %
Genuine	45	Same	-	4
Imposter	30	-	0	-
Genuine	45	Diff	-	15

References

[1] Speaker recognition workshop notebook. *Proc. NIST 1999 speaker recognition workshop, University of Maryland, USA*, Jun 3-4 1999.

[2] B. S. Atal. Automatic speaker recognition based on pitch contours. *J. Acoust. Soc. Amer.*, 52(6):1687–1697, 1972.

[3] B. Yegnanarayana, S.P. Kishore and A.V.N.S. Anjani. Neural network models for capturing probability distribution of training data. In *Int. Conference on Cognitive and Neural Systems*, May 2000.

[4] Sadaoki Furui. Cepstral analysis technique for automatic speaker verification. *IEEE Trans. Acoust., Speech, Signal Processing*, 29(2):254–272, April 1981.

[5] H. Gish and M. Schmidt. Text–independent speaker identification. *IEEE Signal Processing Magazine*, pages 18–32, October 1994.

[6] M. Shajith Ikbal, Hemant Misra, and B. Yegnanarayana. Analysis of autoassociative mapping neural networks. In *Int. Joint Conf. on Neural Networks*, Washington, USA, July 1999.

[7] Mark A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE*, 37:223–243, February 1991.

[8] M. Shajith Ikbal. *Autoassociative neural network models for speaker verification*. MS dissertation, Indian Institute of Technology, Department of Computer Science and Engg., Madras, May 1999.

[9] Douglas O'Shaughnessy. Speaker recognition. *IEEE ASSP Magazine*, pages 4–17, 1986.

[10] Douglas A. Reynolds. Comparision of background normalisation methods for text–independent speaker verification. In *Eurospeech*, pages 963–966, Greece, 1997.

[11] S.P. Kishore and B. Yegnanarayana. Speaker verification: Minimizing the channel effects using autoassociative neural network models. In *Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing*, 2000.

[12] B. Yegnanarayana. *Artificial Neural Networks*. Prentice-Hall of India, New Delhi, 1999.