IJETIR
ISSN (Online) 2583-0554
Impact Factor: 5.731
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 12, December 2025

Limited Data Speaker Verification using Different Modelling Techniques

Jayanthi Kumari T R[1], Anita R[2], Rajesh L[3], Navya V[4]
East Point College of Engineering & Technology, Bangalore, Karnataka, India[1,2,4]
S-Vyasa Deemed University, Bangalore, Karnataka, India[3]
trjayanthikumari.ece@eastpoint.ac.in, anitar.ece@eastpoint.ac.in, dr.rajesh_l@svyasa.edu.in
drvnavya.ece@eastpoint.ac.in

Abstract: *Most modeling techniques in speaker verification (SV) systems rely on sufficient data to effectively model the speaker, resulting in good performance. However, studies have shown that performance significantly degrades when the available training data is limited, due to inadequate speaker modeling. Additionally, insufficient speech data during testing can lead to unreliable verification decisions. Therefore, the objective of this work is to achieve reliable and accurate performance in SV systems even under data-limited conditions. Several modeling techniques have been employed in SV systems, including CVQ, FVQ, SOM, LVQ, GMM, and GMM-UBM*

Keywords: *speaker verification*

I. INTRODUCTION

Modeling techniques aim at generating good representative vectors for features of the speaker. The majority of the modeling techniques uses sufficient data for modeling the speaker and provides good performance in speaker verification (SV) system. In this study, it is reported that when available training data is less, the performance is degrade because of poor speaker modeling. Also, insufficient test speech leads to unreliable decision during testing. Therefore, the objective of the work is to get good and reliable performance in SV system under limited data. Many modeling techniques have been using for SV system. In that few are CVQ, FVQ, SOM, LVQ, GMM, GMM-UBM and i-vector. The success of the modeling techniques depends on different clustering algorithm. Among these modeling techniques, the most frequently used modeling technique is GMM [1]. This is because GMM works well for sufficient data [2]. The literature survey reveals that for speaker recognition under limited data GMM-UBM is the most widely used [2]. The SV using i-vector SV done for sufficient train and limited test data [3][4].

In this work, we studied the performance of different modeling techniques as the working principle of each modeling techniques is different.

II. DATABASE FOR THE STUDY

The NIST-SRE-2003 database consists of speech data from 356 speakers (149 male and 207 female). The spontaneous speech of speakers was collected over cellular phone, sampledat 8 kHz and stored with 16 bits/sample resolution for useas training and testing data. The range of speech data varies from few seconds to few minutes. Since the database is not meant for limited data condition, we have taken four, five and six seconds of each speaker data to create the databasefor the present work. A detailed description of the databasecan be found in the NIST-SRE-2003 plan (NIST 2003) [5]**.**

III. LIMITED DATA SV STUDIES USING DIFFERENT MODELLING TECHNIQUES

The importance of modeling technique is to create model using speaker-specific feature vectors. In this process a set of feature vectors are grouped into their representative
vectors. State-of-the-art uses various modeling techniques for SV system. It contains different types of pattern matching techniques like template matching, probabilistic

model and artificial neural network. The CVQ, FVQ, SOM, LVQ, GMM, GMM-UBM and i-vector modeling techniques are used in the present work

I. CVQ

CVQ is a process of clustering all feature vectors from large vector space into non-overlaping clusters with crisp boundaries. In CVQ, SV depends on measuring distance
between test vectors and the models of the train speakers of the given database [6]. The primary codebook is obtained by splitting training data and initial code vector
is determined as the average of entire training data. The initial code vector again divide into two and this algorithm repeat with these two codebooks. The obtained two codebooks are divided into four codebooks and iterative algorithm is repeated till the required codebook size is achived..

II. FVQ

In case of FVQ, clustering will take place
in such a way that all the feature vectors in feature space are placed into overlapping clusters with fuzzy boundaries' [7]. The feature vectors allotted to all the clusters with different degree of association, as directed by membership function. During training different codebook size is designed with the help of binary split and fuzzy c-means clustering [7]. The nature of clustring purely depends on learning rate parameter and to get better recognition rate the learning rate parameter is tuned properly.

III. SOM

SOM is one of the popular method of clustering, which needs no human intervention, that means unsupervised and provides competitive learning [8, 9]. Still retaining the underlying structure of the input data, SOM reduces dimensionality of the input data and thus provides a visualization of the input data in the low dimension. SOM works in two phases. During training process the values for the input variables are continuously balanced trying to safeguard neighborhood connections that exist inside the input data set and that how the input topology is preserved.

IV. LVQ

The LVQ is a supervised learning that uses a competitive (winner-take-all) learning strategy. In LVQ the class boundaries decided by nearest-neighbor rule and a winner
Takes-it-all paradigm

IV. GAUSSIAN MIXTURE MODEL (GMM)

The Gaussian mixture model (GMM) is a density estimator and is one of the most commonly used types of classifier for speaker recognition [10]. In GMM, the distribution of the feature vector is modelled clearly using a mixture of M Gaussians. Given a collection of training vectors, maximum likelihood model parameters are estimated using iterative expectation-maximization (EM) algorithm [11].The EM algo- rithm iteratively refines the GMM parameters to monotonically increase the likelihood of the estimated model for the observed feature vectors. The complete GMM is parameterized by the mean vectors, covariance matrices and mixture weights from all components densities. These parameters are collectively represented by the notation,

$$\lambda = \{\omega i, \mu i, \Sigma i\} \quad (1)$$

During testing the speaker recognition system uses the matching and decision logic [11]. Test features vectors are matched with the reference models, matching gives a score which represents how well the test feature vectors are close to the reference models. Decision will be taken based on a final set of matched scores, which depends on the threshold value. In GMM, Log likelihood ratio test is used for testing.

V. GMM-UNIVERSAL BACKGROUND MODEL (GMM-UBM)

UBM is a large GMM which represents the speaker inde- pendent distribution of features. UBM is generally built using large population of speech. UBM is the core part of GMM- UBM speaker verification system. UBM should be balanced with respect to male and female speakers. To train a UBM, the simplest approach is to merely pool all the data and use it to train the UBM via the EM algorithm. Maximum a posteriori (MAP) adaptation integrates coupled target and background speaker model components is an effective way of performing speaker recognition [12]. During the testing stage, log likelihood ratio test is used for testing.

VI. RESULTS AND DISCUSSION

In this work, the experiments are conducted using different features for SV system. The modeling techniques used are template and probabilistic. The verification performance of the system can be calculated by using EER. The experiments are conducted using NIST-2003 database. The training/testing data is of duration 3s-3s, 4s-4s, 5s-5s, 6s-6s, 9s-9s and 12s-12s is used.

Table 1: Minimum EER(%) of individual modeling using MFCC and LPCC for limited data.

Train/Test data	MFCC				LPCC			
	CVQ	FVQ	SOM	LVQ	CVQ	FVQ	SOM	LVQ
3s-3s	43.49	43.36	43.40	43.21	40.78	39.42	39.46	39.37
4s-4s	43.27	43.17	43.21	43.23	40.10	37.71	37.90	37.23
5s-5s	42.54	42.23	42.41	41.43	38.84	37.48	37.53	37.18
6s-6s	42.14	42.18	41.86	41.03	38.52	36.35	36.35	36.47
9s-9s	38.34	38.30	38.25	38.19	31.79	29.08	29.26	29.26
12s-12s	35.86	35.16	35.77	35.75	29.53	27.23	27.24	27.24

The same set of experiments is conducted for ΔΔMFCC and ΔΔLPCC features. The least EER of individual modeling for different training/testing data using ΔΔMFCC and ΔΔLPCC were calculated and observations are recorded in Table 2. In case of ΔΔMFCC features, LVQ is having reduced EER which is less by 1.31%, 0.04% and 0.01% of CVQ, FVQ and SOM respectively. Further, the same experiment is carried out for 44LPCC feature, LVQ is having reduced EER which is less by 2.07%, 0.8% and 0.3% of CVQ, FVQ and SOM respectively. The same trend is observed for both feature extraction techniques for remaining data sizes.

From Table 1 and Table 2 we observed that, LPCC and ΔΔLPCC features gives better performance than MFCC and ΔΔMFCC features respectively in case of limited training/testing data. From these experimental results it was observed that when training/testing data increases performance will also increases. Among these four individual modeling techniques, FVQ and LVQ performance is better compared to other two modeling techniques. The reason is, in FVQ each of the feature vectors are related with every one of the clusters, comparatively more features for each cluster associated and therefore code vector might be more valid. Because of this reason FVQ performance is better than CVQ and SOM. In case of SOM, to generate the codevector unsupervised learning method is used.

In SOM, actual distribution of feature vector clustering will take place. Therefore, disturbution of feature vector is inuenced by clustering. Further, the improvement in the performance of LVQ implies that employing supervised learning over initially

Table 2: Minimum EER(%) of individual modeling using ΔΔMFCC and ΔΔLPCC for limited data.

Train/Test data	ΔΔMFCC				ΔΔLPCC			
	CVQ	FVQ	SOM	LVQ	CVQ	FVQ	SOM	LVQ
3s-3s	42.41	41.37	41.46	40.21	39.52	38.25	37.75	37.45
4s-4s	41.05	40.13	41.11	39.10	37.7	36.21	37.17	36.13
5s-5s	40.87	40.10	40.37	38.13	37.54	36.15	36.58	35.77
6s-6s	39.92	38.16	39.29	38.52	36.58	34.73	34.28	35.43
9s-9s	33.42	32.11	32.38	32.14	29.26	27.10	27.77	28.21
12s-12s	31.32	30.26	30.44	30.15	27.28	26.19	26.28	25.54

VII. RESULTS AND DISCUSSION

In this work, we have demonstrated the importance of individual modeling techniques. First, we studied the working principles of CVQ, FVQ, SOM and LVQ modeling techniques. We observed that, performance of FVQ and LVQ is better EER compared to CVQ and SOM in 13 and 39 dimensions. For the short utterance (3s-3s to 5s-5s) GMM-UBM modeling gives better performance.

REFERENCES

[1] D.A. Reynolds, \Speaker identification and verification using Gaussian mixture speaker models," Speech commun., vol. 17, no. 1, pp. 91{108, 1995..

[2] P. Angkititrakul and J.H.L. Hansen, \Discriminative in-set/out-of-set speaker recognition," IEEE Trans. Audio, Speech and Language Process., vol. 15, no. 2, pp. 498{508, 2007.

[3] N. Dehak, P. Dumouchel and P. Kenny, \Modeling prosodic features with joint factor analysis for speaker verification," IEEE Trans. Audio, Speech and Language Process., vol. 15, no. 7, pp. 2095-2103, 2007..

[4] R.K. Das, S. Jelil and S.R.M. Prasanna, \Development of Multi-Level Speech based Person Authentication System," J. Signal Processing Systems, pp. 1-13, 2016.

[5] B.S. Atal, "Automatic recognition of speakers from their voices," Proc. IEEE, vol. 64, no. 4, pp. 460-475, 1976.

[6] G.S. Raja, "Feature analysis and compensation for speaker recognition under stressed condition," Ph.D. dissertation, Indian Institute of Technology Guwahati, Dept. of of Electronics & Communication Engg., Guwahati, India, Jul. 2007.

[7] H.S. Jayanna and S.R.M. Prasanna, "An experimental comparison of modeling techniques for speaker recognition under limited data condition," Sadhana, vol. 34, no. 5, pp. 717-728, 2009.

[8] T. Kohonen, "The self-organizing map," in Proc. IEEE, vol. 78, no. 9, Belgrade, 1990, pp. 1464-1480.

[9] M.L. westerlund, "Classi_cation with Kohonen self-organizing maps," Soft computing, 2005.

[10] B. Yegnanarayana, S.R.M. Prasanna, J.M. Zachariah and C.S. Gupta, \Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system," IEEE Trans. Speech and Audio Process., vol. 13, no. 4, pp. 575-582, 2005.

[11] D. O'shaughnessy, Speech commun.: human and machine. IEEE press, 1987.

[12] J.E. Luck, "Automatic speaker verification using cepstral measurements," J. Acoustical Soc. Amer., vol. 46, no. 4B, pp. 1026-1032, 1969