J Inf Process Syst, Vol.13, No.4, pp.928~940, August 2017
https://doi.org/10.3745/JIPS.02.0025

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Combination of Classifiers Decisions for Multilingual Speaker Identification

B. G. Nagaraja* and H. S. Jayanna**

Abstract
State-of-the-art speaker recognition systems may work better for the English language. However, if the same system is used for recognizing those who speak different languages, the systems may yield a poor performance. In this work, the decisions of a Gaussian mixture model-universal background model (GMM-UBM) and a learning vector quantization (LVQ) are combined to improve the recognition performance of a multilingual speaker identification system. The difference between these classifiers is in their modeling techniques. The former one is based on probabilistic approach and the latter one is based on the fine-tuning of neurons. Since the approaches are different, each modeling technique identifies different sets of speakers for the same database set. Therefore, the decisions of the classifiers may be used to improve the performance. In this study, multitaper mel-frequency cepstral coefficients (MFCCs) are used as the features and the monolingual and cross-lingual speaker identification studies are conducted using NIST-2003 and our own database. The experimental results show that the combined system improves the performance by nearly 10% compared with that of the individual classifier.

Keywords
Classifier Combination, Cross-lingual, Monolingual, Multilingual, Speaker Identification

1. Introduction

Speaker identification aims at recognizing the speaker by their voice [1]. Speaker identification is a one-to-many comparison (i.e., the system identifies a speaker from a database of N known speakers). Depending on the mode of operation, speaker identification can be either text-dependent or text-independent [2]. In the former case, the speaker must speak a given phrase known to the system, which can be fixed or prompted. In the latter case, the system does not know the phrase spoken by the speaker. Speaker identification can be performed in the monolingual and cross-lingual modes [3]. In monolingual speaker identification, training and testing languages for a speaker are the same; whereas, in cross-lingual speaker identification, training is done in one language (say x) and testing is done in a different language (say y).

The spoken language mismatch is one of the factors resulting in performance degradation in multilingual speaker recognition systems [4]. For speaker recognition tasks, numerous speech features and modeling techniques have been proposed over the years [5,6]. However, it is still difficult to

Manuscript received January 27, 2014; first revision July 21, 2014; accepted August 5, 2014; onlinefirst August 10, 2015.
Corresponding Author: Nagaraja B. G. (nagarajbg@gmail.com)
* Dept. of E&CE, Jain Institute of Technology, Davangere, Karnataka, India (nagarajbg@gmail.com)
** Dept. of IS&E, Siddaganga Institute of Technology, Tumkur, Karnataka, India (jayannahs@gmail.com)

implement a single classifier that exhibits sufficiently high performance in a multilingual environment. The classifier fusion has received significant attention in the recent years.

There are two ways of combining the classifiers by using a serial combination or parallel combination [7]. In serial combination, classifiers are sequentially arranged and the result from the previous classifier is fed to the next classifier. Parallel combination organizes the classifiers in parallel [7]. The system performance in a parallel combination depends on the combination function. Ideally, the combination function must take advantage of the strengths of all the classifiers, avoid their weaknesses, and improve classification accuracy [8]. The key constraint for a combination function that uses the output of the individual classifiers is that the classifiers should not have the same opinion with each other when they misclassify the pattern [9]. The Gaussian mixture model-universal background model (GMM-UBM) and learning vector quantization (LVQ) classifiers are different with respect to their working principle. Hence, they may be combined to further improve the performance of a multilingual speaker recognition system.

In this work, text-independent monolingual and cross-lingual speaker identification studies are conducted using multitaper mel-frequency cepstral coefficient (MFCC) features, GMM-UBM, and LVQ. The remainder of this paper is organized as follows: in Section 2, a brief overview of classifiers combination techniques for speaker recognition is presented. Section 3 describes the speech database used for the study. Feature extraction using multitaper MFCC and speaker modeling using GMM-UBM and LVQ are presented in Section 4. Section 5 describes monolingual speaker identification using the multitaper MFCC and GMM-UBM methods. Monolingual speaker identification using multitaper MFCC and LVQ technique is presented in Section 6. The combination of classifiers for a monolingual speaker identification task is given in Section 7. Discussions on the cross-lingual experimental results on our own database are presented in Section 8. Section 9 gives the summary and conclusions of this study.

2. Related Work

In [10], a hybrid Karhunen-Loeve transform and GMM approach based on two-stage classifiers has been proposed for text-independent speaker identification. The experimental results on the 500 Mandarin speakers showed that the combination scheme is helpful to both classification accuracy and computational cost. Masho and Skosan [6], have combined the decisions of two systems for the speaker recognition task. One system was based on the MFCC features and the other on the parametric feature sets algorithm. The combined classifier system produced a good speaker identification rate on the NTIMIT database. In [11], a new classifier combination method based on signal strength was proposed to support the decision-making process. Based on various real-world machine learning data sets, the proposed method showed better results compared with the existing voting strategies and margin-based classifiers.

In [12], the scores from the GMM approach, support vector machine, and decision tree classifiers are combined for the text-independent speaker identification task. The experimental results on dialect DR1 (47 speakers) of the TIMIT corpus showed that the combined classifier outperforms the individual classifiers. A sparse regularized logistic regression score fusion method for speaker verification was proposed in [13]. The proposed method was evaluated using the NIST SRE2010 corpus. The experimental

results showed that the sparse regularization achieved improvement overan un-regularized variant, except in a telephony-telephony speech data condition. In our previous work, the significance of combining the evidence from multitaper MFCC and linear prediction residual features for multilingual speaker identification with the constraint of the limited data condition was studied. The experimental results showed that the combined evidence improves the performance by nearly 8%–10% compared to individual evidence [14].

3. Speech Database for the Study

The monolingual (English language) speaker identification studies were conducted on 30 randomly selected speakers (17 male and 13 female) from the NIST-2003 database [15]. Since the standard multilingual database was not available, multilingual experiments were carried out on our own database of 50 speakers that was created from the speakers who speak three different languages (E-English, H-Hindi, and K-Kannada). This database includes 30 male and 20 female speakers. The voice recording was done in an engineering college laboratory. The speakers were undergraduate students and faculty members in an engineering college. The age of the speakers varied from 18–35 years. The speakers were asked to read small stories in three different languages. The training and testing data was recorded in different sessions with a minimum gap of two days. The approximate training and testing data length is two minutes. Recording was done using free downloadable wave surfer 1.8.8p3 software and the Beetel headphone-250 with a frequency range of 20–20 kHz. The speech files were stored in a.wav format.

4. Feature Extraction and Modeling

4.1 Multitaper MFCC

Let $F=(f[0],f[1],...,f[N-1])^T$ denote one frame of speech (N samples) signal. The multitaper MFCCs were calculated for 20 ms speech segments with 50% overlapping (frame rate). For the multitaper MFCC methods, the spectrum $S(f)$ is obtained by [16].

$$S(f)=\sum_{j=1}^{K}\lambda(j)\left|\sum_{n=0}^{N-1}w_j[n]f[n]e^{-i2\pi fn/N}\right|^2 \quad (1)$$

Here K represents the number of multitapers used. $W_j=(w_j[0],w_j[1]...,w_j[N-1])^T$ are the multitaper weights and $j=1,2,..,K,$ are used with the corresponding weights $\lambda(j)$. A signal flow diagram of the multitaper spectrum estimator is illustrated in Fig. 1. Fig. 2 shows the block diagram representation of the multitaper MFCC method.

We only considered the first 13 dimensional feature vectors (excluding the 0^{th} coefficient) computed using 22 filters in the filter bank. Cepstral mean subtraction was applied to the multitaper MFCC to remove the linear channel effect. Silence and low-energy speech parts were removed using an energy-based voice activity detection technique [17]. The threshold we used was 0.06 times the average frame

energy for the selection of speech frames.

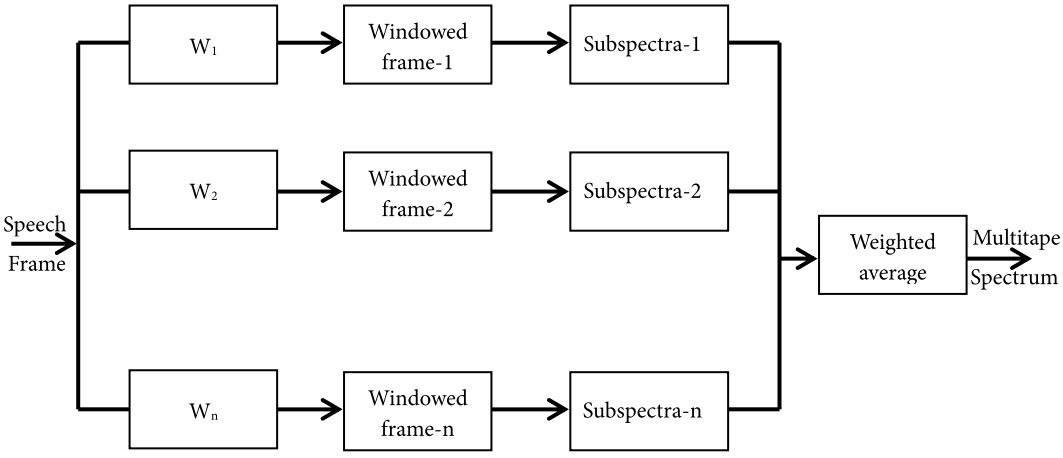

Fig. 1. Signal flow diagram of the multitaper spectrum estimation.

Different types of tapers have been proposed for spectrum estimation in [16]. It was mentioned that the range of K should be between 3 and 8 and also recommended to start with $K = 6$. In this work, sine-weighted cepstrum estimators (SWCE) [18], Thomson [19], and Multipeak [20] multi-tapers were used with $K = 6$, 7, and 8 windows.

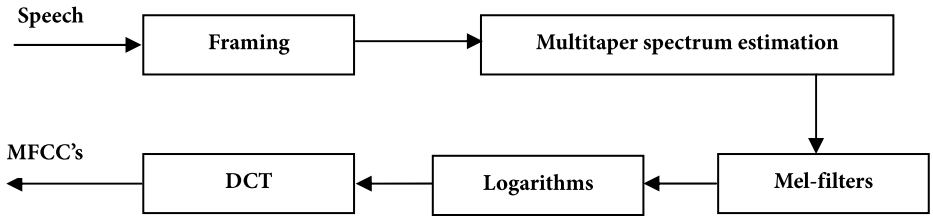

Fig. 2. Block diagram of multitapermel-frequency cepstral coefficient (MFCC) features extraction technique.

4.2 Speaker Modeling Using GMM-UBM

The GMM-UBM is the most widely used probabilistic modeling technique in speaker recognition [21,22]. For building the UBM, we used approximately two hours of speech data from all of the 138 speakers of the YOHO database (the first 18 speech files from the enroll data) [23]. Adapting only the mean vectors of the UBM using maximum a posteriori adaptation algorithm created the gender independent speaker specific models. The parameters of the GMM (mean vector, covariance matrix, and mixture weights) were estimated using the expectation maximization algorithm. We modeled speakers with 16, 32, 64, 128, and 256 Gaussian mixtures.

4.3 Speaker Modeling Using LVQ

LVQ is a supervised version of vector quantization. LVQ algorithms directly describe class boundaries based on the nearest-neighbor rule and a winner-takes-it-all paradigm [24]. If the class label of the input vector and the code vector agree, then the code vector is moved in the direction of the input vector. Otherwise, the code vector is moved away from the input vector. Suppose X_t is an input vector at

time t, and W_j is the weight vector for class j at time t. Let ζ_{wc} denote the class associated with the weight vector W_c and ζ_x denote the class label of the input vector X. The weight vector W_c is adjusted as follows [22,25]:

1. if $\zeta_{wc} = \zeta_x$, then:
   $W_c(t+1) = W_c(t) + \eta(t)[x - W_c(t)]$; where $0 < \eta(t) < 1$, η = learning rate
2. else:
   $W_c(t+1) = W_c(t) - \eta(t)[x - W_c(t)]$
3. The other weight vectors are not modified.

5. Monolingual Speaker Identification Using Multitaper MFCC and GMM-UBM

An effect of the choice of multitaper type and the number of tapers on speaker identification for 30 randomly selected speakers (20 seconds of training and test data) for the NIST-2003 and our own databases using the GMM-UBM classifier is shown in Figs. 3 and 4. It was observed that the speaker identification system gives the highest performance of 50% and 90% using the SWCE multitaper for K = 6 windows for NIST-2003 and our own database, respectively. Furthermore, it was observed that the SWCE multitaper performs better than the Thomson and Multipeak multitaper techniques. Henceforth, we used the SWCE multitaper MFCC (K = 6) as features for all of our experimental studies.

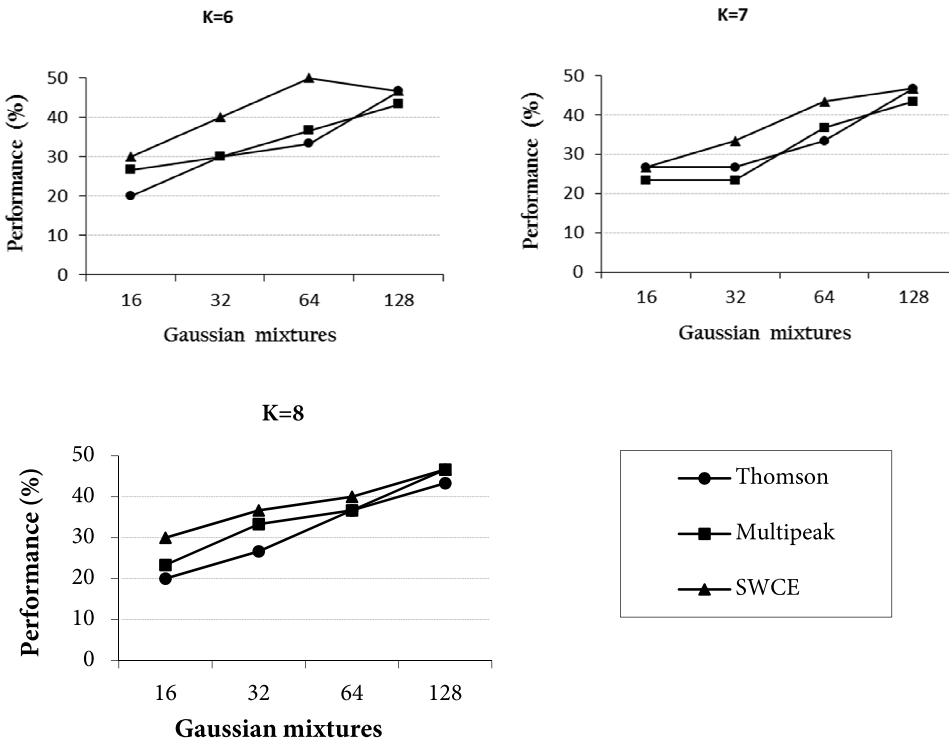

Fig. 3. Speaker identification performance (%) for randomly selected 30 speakers of NIST-2003 database using different multitapers for K = 6, 7, and 8. SWCE=sine-weighted cepstrum estimator.

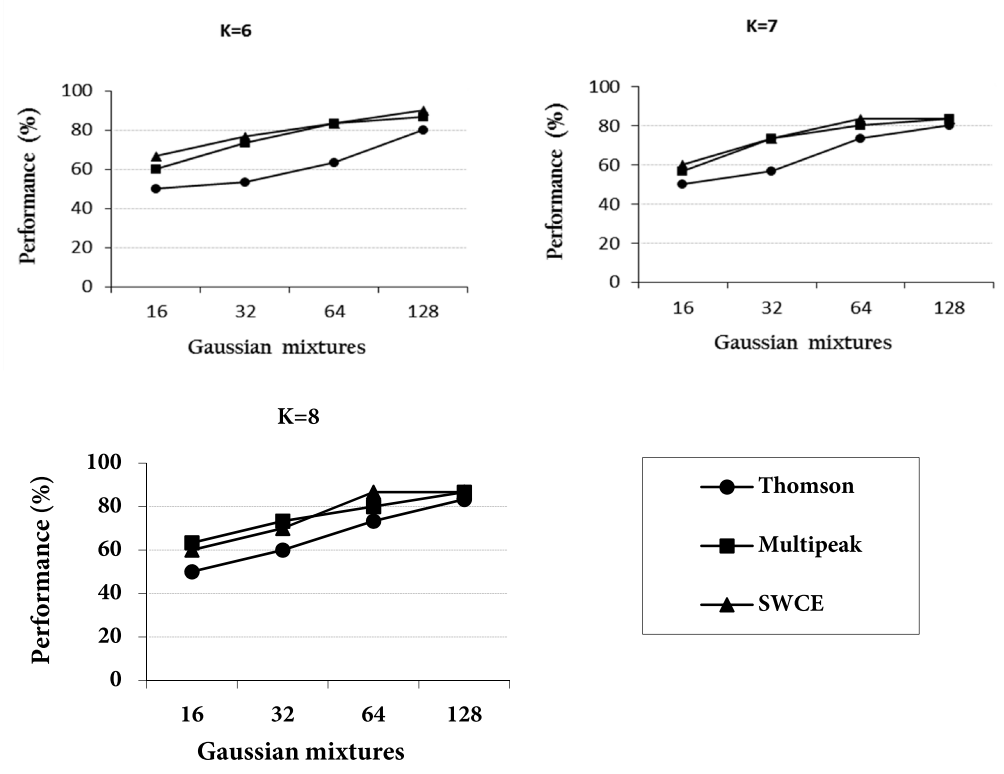

Fig. 4. Speaker identification performance (%) for randomly selected 30 speakers of our own database using different multitapers for K = 6, 7, and 8. SWCE=sine-weighted cepstrum estimator.

6. Monolingual Speaker Identification Using Multitaper MFCC and LVQ

The LVQ performance depends on the parameters like η and iterations. The identification performance for the 30 randomly selected speakers (20 seconds of training and test data) of the NIST-2003 and our own databases and different η and iterations are given in Tables 1 and 2. The speaker identification system gives the highest performance of 53.33% (η = 0.02 and iterations = 650 × codebook size [CS]) and 86.66% (η = 0.03 and iterations = 500 × CS) for the NIST-2003 and our own database, respectively. Though the performance of LVQ is lesser than GMM-UBM, the combination of classifiers may improve the performance.

Table 1. Speaker identification performance (%) for randomly selected 30 speakers of NIST-2003 database using LVQ technique

Iterations	η	Codebook size (CS)			
		16	32	64	128
500 × CS	0.01	30.00	36.66	43.33	46.66
500 × CS	0.02	23.33	36.66	50.00	43.33
500 × CS	0.03	26.66	33.33	46.66	40.00
600 × CS	0.02	30.00	40.00	50.00	46.66
650 × CS	0.02	23.33	33.33	46.66	**53.33**
700 × CS	0.02	26.66	30.00	36.66	46.66

Table 2. Speaker identification performance (%) for randomly selected 30 speakers of our own database using LVQ technique

Iterations	η	Codebook size (CS)			
		16	32	64	128
500 × CS	0.01	53.33	60.00	73.33	83.33
500 × CS	0.02	46.66	56.66	83.33	76.66
500 × CS	0.03	50.00	66.66	76.66	**86.66**
500 × CS	0.04	53.33	56.66	80.00	76.66
600 × CS	0.03	50.00	60.00	66.66	80.00
700 × CS	0.03	53.33	63.33	70.00	76.66

7. Combination of Classifiers for Speaker Identification

The problem of combining classifiers that use different pattern representations was studied in [9]. They also provided a common theoretical framework for combining classifiers. In our work, the frame scores obtained from GMM-UBM (classifier-1) and LVQ (classifier-2) classifiers are not directly usable because of the incompatibility of their scales. Hence, for each speaker, the confidence score (C_i) for a given test signal is computed as:

$$C_i = \frac{S_i}{\max_{j=1}^{N} S_j}, \tag{2}$$

where S_i is the frame score of each speaker and N is the total number of enrolled speakers. Let $C_i^{(1)}$ refer to the confidence scores associated with classifier-1 and $C_i^{(2)}$ corresponding to the confidence scores associated with classifier-2. For combining classifiers, we can use the Kittler, Hatef, Duin, and Mataz (KHDM) rules, which are as follows [6,9]:

$$\text{Sum rule; } S_{sum}(C) = \arg\max_{s=1}^{N}\left[\sum_{i=1}^{2} C_s^{(i)}\right] \tag{3}$$

$$\text{Product rule; } S_{prod}(C) = \arg\max_{s=1}^{N}\left[\prod_{i=1}^{2} C_s^{(i)}\right] \tag{4}$$

$$\text{Maximum rule; } S_{\max}(C) = \arg\max_{s=1}^{N}\left[\max_{i=1}^{2} | C_s^{(i)} |\right] \tag{5}$$

$$\text{Minimum rule; } S_{\min}(C) = \arg\max_{s=1}^{N}\left[\min_{i=1}^{2} | C_s^{(i)} |\right] \tag{6}$$

In this work, a parallel multiple classifier architecture with frame score level base classifiers were combined using KHDM rules. The best identification performance of individual classifiers an combined classifier using KHDM rules for 30 randomly selected speakers (20 seconds of training and test data) of the NIST-2003 and our own databases are given in Table 3.

Table 3. Applying KHDM rules on the classifiers

Modeling technique	Identification performance (%)	
	NIST-2003 database	our own database
GMM-UBM	50.00	90.00
LVQ	53.33	86.66
GMM-UBM-LVQ (Sum)	60.00	**96.66**
GMM-UBM-LVQ (Prod)	53.33	93.33
GMM-UBM-LVQ (Max)	53.33	90.00
GMM-UBM-LVQ (Min)	56.66	93.33

GMM-UBM=Gaussian mixture model-universal background model, LVQ=learning vector quantization.

Table 4. Number of speakers identified by the GMM-UBM, LVQ and combined GMM-UBM-LVQ using sum rule for 30 speakers of our own database (√, identified; x, not identified)

Speaker	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
GMM-UBM	√	√	√	**x**	√	√	√	√	√	√	**x**	√	√	√	√
LVQ	√	√	√	**x**	√	√	√	√	√	√	√	√	√	√	√
Combined	√	√	√	**x**	√	√	√	√	√	√	√	√	√	√	√

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	No.
√	√	√	√	√	√	√	√	√	√	√	**x**	√	√	√	**27**
x	√	√	**x**	√	√	√	√	**x**	√	√	√	√	√	√	**26**
√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	**29**

GMM-UBM=Gaussian mixture model-universal background model, LVQ=learning vector quantization.

The proposed combined system yields a good identification rate in all of the speaker identification experiments. Table 3 shows that the performance of the LVQ system is almost the same as that of the GMM-UBM. However, speakers identified by the GMM-UBM and LVQ systems are different and are shown in Table 4. The speaker numbers 11 and 27 are not identified by GMM-UBM but are identified by LVQ. Similarly, speaker numbers 16, 19, and 24 are not identified by LVQ but are identified by GMM-UBM. The combined system identified all speakers, except speaker number four. The same trend was also observed for the NIST-2003 database. The improvement in performance may be due to the employment of a different working principle in GMM-UBM and LVQ. The LVQ modeling technique is based on a non-parametric approach, whereas, GMM-UBM is based on a parametric approach. Hence, this combination gives the best identification performance [22]. The sum rule outperformed the other combination schemes since it is less sensitive to estimation errors [6,9]. Henceforth, we used the sum rule for all of the experimental studies.

To verify the robustness of the proposed method for a large set of speakers and for different languages (English, Hindi, and Kannada), we conducted the experiments using 50 speakers from our own database using three different languages. Note: *x*/*y* indicates training with language *x* and testing with language *y* (e.g., *E*/*K* indicates training with the English language and testing with the Kannada language). The monolingual experimental results for the 50 speakers from our own database for 20 seconds of training and testing data and for different codebook sizes/Gaussian mixtures are given in Fig. 5.

The speaker identification system trained and tested with the English language (E/E) gave the highest performance of 84% and 90% for LVQ and GMM-UBM classifiers, respectively. The performance of the

speaker identification system trained and tested with the Hindi language (H/H) was 78% and 84% for LVQ and GMM-UBM classifiers, respectively. The speaker identification system trained and tested with the Kannada language (K/K) gave the highest performance of 80% for LVQ and GMM-UBM classifiers.

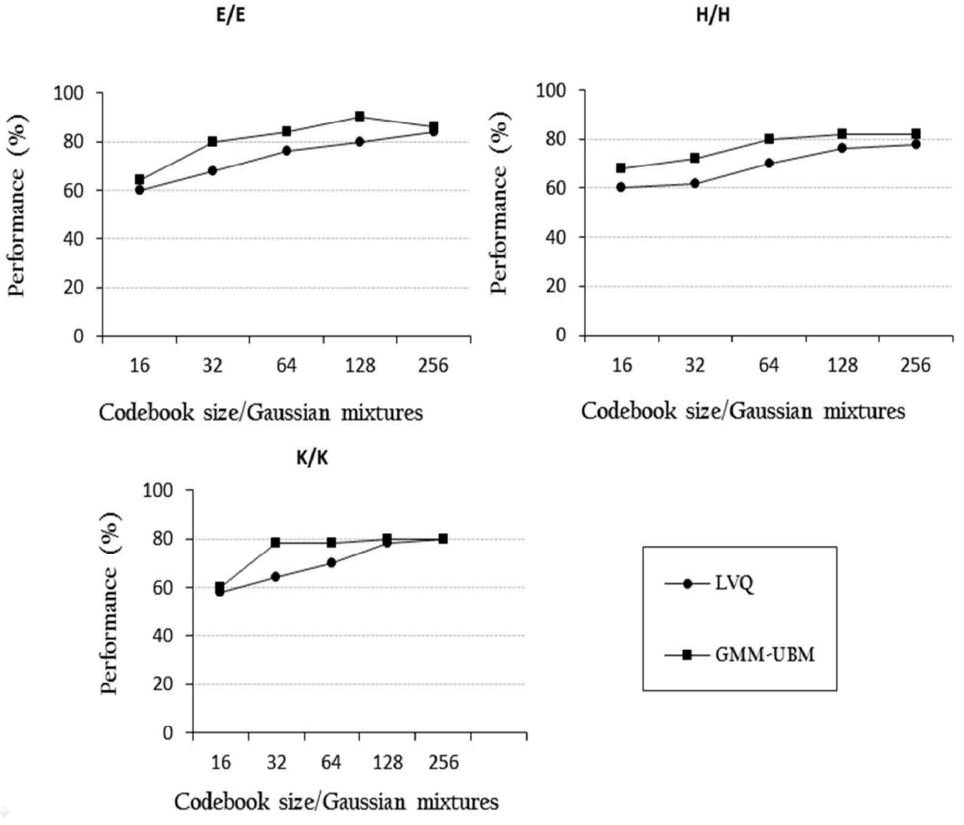

Fig. 5. Performance of monolingual speaker identification for 50 speakers of our own database using 20 seconds of training and testing data. GMM-UBM=Gaussian mixture model-universal background model, LVQ=learning vector quantization.

Table 5 shows the best performance (%) of combined classifiersfor monolingual speaker identification for the 50 speakers from our own database using the sum rule. It was observed that the performance of a combined classifier system is better than the individual classifiers.

Table 5. Applying a sum rule on the classifiers for monolingual speaker identification using 50 speakers of our own database

Train/test language	Identification performance (%)
E/E	94.00
H/H	90.00
K/K	90.00

E/E=English language, H/H=Hindi language, K/K=Kannada language.

8. Cross-lingual Speaker Identification

In the previous section we demonstrated the usefulness of combining the two classifiers for monolingual speaker identification. In this section, cross-lingual studies are conductedusing the proposed combined classifier technique. Since the data was collected in three different languages to study the robustness of the system, the experiments were conducted for six cases (H/E, K/E, E/H, K/H, E/K, and H/K). The cross-lingual experimental results for the 50 speakers from our own database for 20 seconds of training and testing data and for different codebook sizes/Gaussian mixtures are shown in Fig. 6. It was observed that the results are better for monolingual experiments than cross-lingual ones. This may be due to the variation in fluency and word stress when the same speaker speaks different languages and may also be due to different phonetic and prosodic patterns of the languages [26].

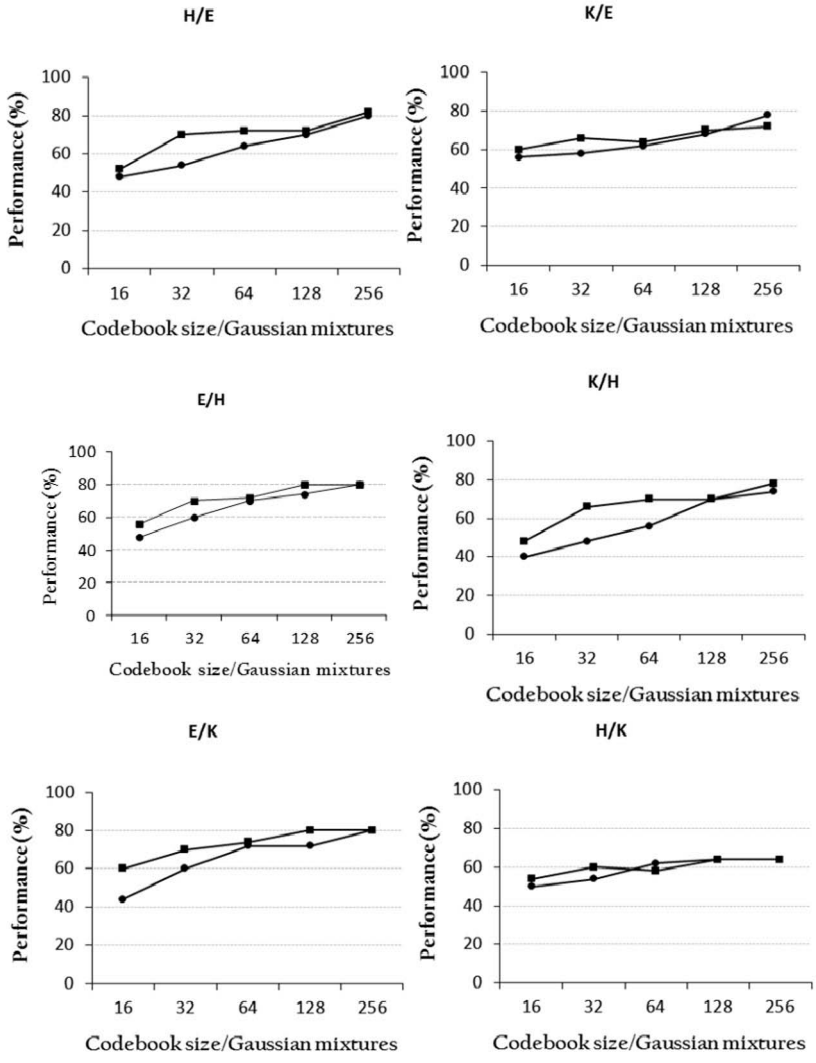

Fig. 6. Performance of cross-lingual speaker identification for 50 speakers of our database using 20 seconds of training and testing data. E/E=English language, H/H=Hindi language, K/K=Kannada language.

Table 6. Applying a sum rule on the classifiers for cross-lingual speaker identification using 50 speakers of our own database

Train/test language	Identification performance (%)
H/E	92.00
K/E	92.00
E/H	88.00
K/H	88.00
E/K	88.00
H/K	84.00

E /E=English language, H/H=Hindi language, K/K=Kannada language.

Table 6 shows the best performance (%) of combined classifiers for cross-lingual speaker identification for the 50 speakers from our own database using the sum rule. Though the performance of a combined classifier system is better than the individual classifiers, the rate of improvements in the identification performance of the proposed combined system is significantly higher for cross-lingual than monolingual. In the monolingual case, the performance of the individual classifiers is sufficiently high and hence, the combined classifier improvement may be less.

9. Conclusions

The combination of GMM-UBM and LVQ based classifiers was studied for monolingual and cross-lingual speaker identification. The results indicated that the proposed combined system can be used for improving the performance of multilingual speaker identification. The performance of various classifier combination methods (sum rule, product rule, maximum rule, and minimum rule) was compared for the multilingual speaker identification task and it was observed that the sum rule outperformed other classifier combination methods. Furthermore, we observed that the performance of the combined system is significantly higher for cross-lingual than for monolingual speaker identification.

References

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

[2] D. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72-83, 1995.

[3] P. H. Arjun, "Speaker recognition in indian languages: a feature based approach," Ph.D. dissertation, Indian Institute of Technology Kharagpur, India, 2005.

[4] M. Akbacak and J. H. Hansen, "Language normalization for bilingual speaker recognition systems," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007),* Honolulu, HI, 2007, pp. 257-260.

[5] G. R. Doddington, M. A. Przybocki, A. F. Martin, and D. A. Reynolds, "The NIST speaker recognition evaluation–overview, methodology, systems, results, perspective," *Speech Communication*, vol. 31, no. 2, pp. 225-254, 2000.

[6] D. J. Mashao and M. Skosan, "Combining classifier decisions for robust speaker identification," *Pattern Recognition*, vol. 39, no. 1, pp. 147-155, 2006.

[7] E. Kim, W. Kim, and Y. Lee, "Combination of multiple classifiers for the customer's purchase behavior prediction," *Decision Support Systems*, vol. 34, no. 2, pp. 167-175, 2003.

[8] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66-75, 1994.

[9] C. C. T. Chen, C. T. Chen, and C. K. Hou, "Speaker identification using hybrid Karhunen–Loeve transform and Gaussian mixture model approach," *Pattern Recognition*, vol. 37, no. 5, pp. 1073-1075, 2004.

[10] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226-239, 1998.

[11] H. He and Y. Cao, "SSC: a classifier combination method based on signal strength," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 7, pp. 1100-1117, 2012.

[12] S. Z. Boujelbene, D. Ben AyedMezghani, and N. Ellouze, "Application of combining classifiers for text-independent speaker identification," in *Proceedings of the 16th IEEE International Conference on Electronics, Circuits, and Systems (ICECS 2009)*, Yasmine Hammamet, 2009, pp. 723-726.

[13] V. Hautamaki, T. Kinnunen, F. Sedlák, K. A. Lee, B. Ma, and H. Li, "Sparse classifier fusion for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1622-1631, 2013.

[14] B. G. Nagaraja and H. S. Jayanna, "Multilingual speaker identification by combining evidence from LPR and multitaper MFCC," *Journal of Intelligent Systems*, vol. 22, no. 3, pp. 241-251, 2013.

[15] The NIST Year 2003 speaker recognition evaluation plan [Online]. Available: http://www.itl.nist.gov/iad/mig/tests/sre/2003/2003-spkrec-evalplan-v2.2.pdf.

[16] T. Kinnunen, R. Saeidi, F. Sedlák, K. A. Lee, J. Sandberg, M. Hansson-Sandsten, and H. Li, "Low-variance multitaper MFCC features: a case study in robust speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 1990-2001, 2012.

[17] J. R. Deller, J. H L Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. New York, NY: Institute of Electrical and Electronics Engineers, 1993.

[18] K. S. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 188-195, 1995.

[19] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055-1096, 1982.

[20] M. Hansson and G. Salomonsson, "A multiple window method for estimation of peaked spectra," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 778-781, 1997.

[21] D. A. Reynolds, "Universal background models," in *Encyclopedia of Biometrics*. Heidelberg: Springer, 2009, pp. 1349-1352.

[22] H. S. Jayanna, "Limited data speaker recognition," Ph.D. dissertation, Indian Institute of Technology Guwahati, India, 2009.

[23] J. P. Campbell Jr, "Testing with the YOHO CD-ROM voice verification corpus," in *Proceedings of 1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'95)*, Detroit, MI, 1995, pp. 341-344.

[24] T. E. F. Filho, R. O. Messina, and E. F. Cabral Jr, "Learning vector quantization in text-independent automatic speaker recognition," in *Proceedings of Vth Brazilian Symposium on Neural Networks*, Belo Horizonte, Brazil, 1998, pp. 135-139.

[25] S. S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York, NY: Prentice-Hall, 1999.

[26] G. Durou, "Multilingual text-independent speaker identification," in *Proceedings of the Multi-Lingual Interoperability in Speech Technology (MIST) Workshop,* Leusden, The Netherlands, 1999.

B. G. Nagaraja http://orcid.org/0000-0002-4702-4953

He received the B.E. degree in Electronics and Communications Engineering from Bapuji Institute of Technology, Davangere, Karnataka, India, in 2004, M.Tech. degree in Computer Science and Engineering from East-West Institute of Technology, Bangalore, India, in 2009 and Ph.D. in Electronics and Communications Engineering from Visvesvaraya Technological University, Belgaum, Karnataka, India, in 2014. He is an Associate Professor at the Department of Electronics & Communication Engineering, Jain Institute of Technology, Davangere, Karnataka, India. His research interests include speech, speaker recognition and multilingual speaker recognition.

H. S. Jayanna http://orcid.org/0000-0002-4342-9339

He received the B.E. degree in Instrumentation and Electronics Engineering from Dr. Ambedkar Institute of Technology, Bangalore University, Bangalore, India, in 1992, the M.E. degree in Electronics from University Visvesvaraya College of Engineering, Bangalore, India, in 1995 and Ph.D. in Electronics and Communication Engineering from the prestigious Indian Institute of Technology, Guwahati, India, in 2009. He has published a number of papers in various national and international journals and conferences apart from guiding a number of UG, PG and research scholars. Currently, he is working as a Professor and Head of the Department of Information Science and Engineering, Siddaganga Institute of Technology, Tumkur, Karnataka, India. His research interests are in the areas of speech, limited data speaker recognition, image processing, computer networks and computer architecture.