Combination of Features for Multilingual Speaker Identification with the Constraint of Limited Data

Nagaraja B.G.
Department of Information Science & Engineering
Siddaganga Institute of Technology
Tumkur-572103, Karnataka, India

H.S. Jayanna
Department of Information Science & Engineering
Siddaganga Institute of Technology
Tumkur-572103, Karnataka, India

ABSTRACT

In the modern day digital automated world, speaker identification system plays a very important role in the field of fast growing internet based communications/transactions. In this paper, speaker identification in the context of mono, cross and multilingual are demonstrated using the two different feature extraction techniques, i.e., Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC) with the constraint of limited data. The languages considered for the study are English (international language), Hindi (national language) and Kannada (regional language). Since the standard multilingual database is not available, experiments are carried out on our own created database of 30 speakers in the college laboratory environment who can speak the three different languages. In case of limited data condition, owing to less data the existing techniques in each stage may not provide good performance. To alleviate the problem of limited data, the vocal tract feature extracted from MFCC and LPCC techniques are combined. As a result the combination of features gives nearly 30% higher performance compared to the individual features for a set of 30 speakers.

Keywords:

Speaker identification, monolingual, crosslingual, multilingual, MFCC, LPCC, VQ

1. INTRODUCTION

Speaker identification is defined as a task of recognizing speakers from their speech [1]. Depending on the mode of operation, speaker identification can be classified as text-dependent identification and text-independent identification [2]. The text-dependent identification requires the speaker to produce speech for the same text, both during training and testing whereas the text-independent identification does not rely on a specific text being spoken [3]. Text-independent speaker identification systems are more versatile, but their accuracy is considerably lower than that of text-dependent systems [4]. To achieve acceptable results in this case, more speech data is usually necessary for both training and testing purposes [4]. Speaker recognition in limited data condition aims at recognizing speaker with the constraint that both training and testing data are limited [5].

In India, more than 50 languages are officially recognized by the Govt. of India and the Indian citizens can speak more than one language fluently. Therefore, the development of multilingual speaker Identification system is a challenging task. Criminals often switch over to another language, especially after committing a crime. So, training a person's voice in one language and identifying him in some other language or in a multilingual environment is of interesting task especially in the Indian context [6]. As we mentioned earlier, in India people have been trained themselves to speak in many languages, for example, when a person from West Bengal comes and settles in Karnataka, he will be knowing English, Hindi, Bengali and Kannada. This advantage can be utilized in machine learning to build a robust speaker recognition system. However, nowadays we cannot ask people to give data for a long period of time as the sufficient speaker recognition system expects. Further, due to increase in the use of communication and internet services for speech mode applications, it is desirable to work with limited data. Speaker recognition under limited conditions could be used in the following applications [6] [7] :

. Access control and transaction authentication - voice dialing, banking transactions through telephone network, teleshopping, database access service, reservation systems, voice mail, law enforcement, activity area restriction.
. Criminal investigation and security access control, in forensic application also the data available may be limited which may be recorded during casual conversation or by tapping the telephone channel.
. To locate the segment of given speaker in an audio stream such as teleconference or meetings. Such data segments usually contain short utterances whose speaker needs to be identified.
. Remote biometric person authentication for electronic transactions where speech is the most preferred biometric feature.

The main parameter of the speech recognition system is the speech or the voice data of the speaker. Characteristics in the speech signal can be attributed to the dimensions of the vocal tract system, pitch, sound decibel level, characteristics of excitation and the learning habits of the speakers [8]. The most effective features for speaker recognition have been the MFCC and LPCC. These features can accurately characterize the vocal tract configuration of a speaker and can achieve good performance [8]. Combination of different features has been proved to be a good method for improving performance in speaker recognition [9].

Xia Wang *et al.* [10] did extensive work on embedded multilingual speech recognition system for Mandarin, Cantonese and English languages and proposes a small foot print, speaker independent multilingual system for isolated word recognitions of the above 3 languages. By sharing phonemes, the memory and computational complexity was reduced by about 40%. They mainly concentrated on the western languages. Ulrike Halsband [11] worked on the bilingual and multilingual language processing and mainly concentrated on the European languages and produced excellent speaker identification results.

Text-dependent multilingual speaker identification for Indian languages using artificial neural networks was carried out by Rajesh Ranjan and his group in [12]. In this work attempt has been made to develop speaker identification system which is used to determine the identity of an unknown speaker among several speakers of known speech characteristics, from a sample of his or her voice. A novel multilingual text independent based speaker identification algorithm was proposed by Geoffrey Duron in [13]. In his paper, he investigated two facets of speaker recognition: cross-language speaker identification and the same language non-native text independent speaker identification. His results indicated that how speaker identification performance would be affected when speakers do not use the same language during the training and testing or when the population is composed of native speakers.

Prateek *et al.* [14] developed a novel algorithm for multilingual speaker recognition using neural networks with the back propagation concepts. They simulated a model of ANN based multilingual speaker recognition system for eight Indian languages (Hindi, English, Assami, Telugu, Punjabi, Rajasthani, Marathi and Bengali) and achieved a success rate of 95%. Bipul Pandey *et al.* [4] proposed a multilingual speaker recognition scheme using adaptive neuro fuzzy inference scheme (ANFIS) for the identification of the speaker and the words spoken. A robust and sufficiently efficient recognition system was tried by them using the features obtained from several sources including the textual and image sources, which produced excellent results. The experimental results show the system to be amply efficient and successful in the recognition of the tasks that are involved.

Zhi-Yi Li *et al.* [9] proposed an effective multi-feature combination in speaker recognition. In their experiments, they used the popular short-term spectral MFCC and spectro-temporal time frequency cepstrum (TFC) to do feature combination followed by Linear discriminant analysis (LDA) and feature-domain latent factor analysis (FDLFA) for channel compensation respectively. The experimental results were carried out on the NIST SRE 2008 short2 telephone-short3 telephone test. Danoush Hosseinzadeh and Sridhar Krishnan [15] combined the Vocal Source and MFCC features for enhanced speaker recognition performance using GMM's. In their work, the 7 spectral features were extracted from the speech spectrum and used to enhance the performance of MFCC-based features. Experiments were carried out on a TIMIT database using text-independent cohort Gaussian mixture model (GMM) speaker Identification system.

Martine Adda Decker discussed the multilingual interoperability concepts in automatic speech recognition and showed that a large number of languages can be considered in automatic speaker identification [16]. Olli Viikki *et al.* investigated the technical challenges that are faced when making a transition from the speaker-dependent to speaker independent speech recognition technology in mobile communication devices. Due to globalization as well as the international nature of the markets and the future applications, speaker independence implies the development and use of language independent automatic speaker recognition to avoid logistic difficulties. Hence, they proposed architecture for embedded multilingual speech recognition systems [17].

Rama Murty and Yegnanarayana [8] combined the evidences from the residual phase and MFCC methods used for speaker recognition and obtained very good results. In their work, the complementary nature of speaker-specific information present in the residual phase in comparison with the information present in the MFCCs was presented.

An attempt was made to recognize multilingual speaker in [7]. In this work, training data of 60 seconds and for different testing data of 1, 3, 7, 10 and 15 seconds are considered for mono and crosslingual experiments. Also, a polynomial classifier of 2^{nd} order approximation was built for speaker modeling. Recently, some attempts have been made to identify the speakers under limited data condition using the concept of Universal Background Model (UBM) to mitigate the sparseness, which requires additional speech data to train the Gaussian mixture model-Universal Background Model (GMM-UBM) [2].

In our previous work [18], an attempt was made to identify speaker in the context of mono and crosslingual speaker identification with the constraint of limited data using MFCC as feature vectors and Vector Quantization (VQ) as modeling technique [18]. We observed that speaker identification system with English language provides good performance in monolingual study. Further, it was observed in crosslingual study that the use of English language either in training or testing gives better identification performance. In this paper, the significance of two feature extraction techniques MFCC and LPCC are demonstrated in the context of mono, cross and multilingual speaker identification with the constraint of limited data.

State-of-the-art speaker recognition uses more than one minutes of speech data. In the present work, sufficient data is used to symbolize the case of having speech data of few minutes ($\geq$ one minute). Alternatively, limited data symbolizes the case of having speech data of few seconds ($\leq$ 15 seconds) [18] [19]. Since the amount of data is small in limited data condition, any one feature extraction technique may not provide enough features for modeling and testing [5]. In this work, first, features are extracted using MFCC and LPCC and then these features are combined. The modeling technique used was Vector Quantization (VQ). Finally, the speaker models are tested against the testing data of respective extraction techniques. The general block diagram of the proposed speaker identification system is shown in Fig. 1.

The paper is organized as follows: Section 2 describes the database used for the experiments. Feature extraction of the recorded speech using MFCC and LPCC methods and modeling using VQ techniques along with the proposed method of the combined combination (MFCC + LPCC) presented in Section 3. In Section 4, monolingual speaker identification is presented. The crosslingual speaker identification is presented in Section 5. In Section 6, multilingual speaker identification scheme is presented. Section 7 gives summary of the present work and scope for the future work.

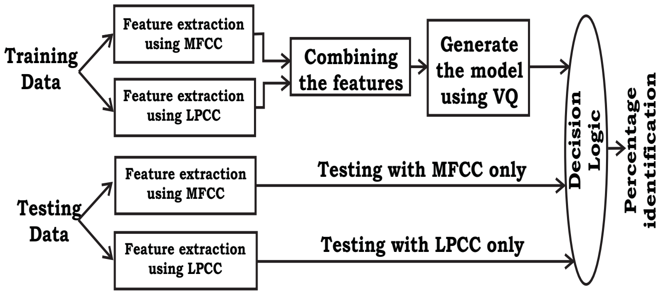

Training Data
Feature extraction using MFCC
Feature extraction using LPCC
Combining the features
Generate the model using VQ
Decision Logic
Percentage identification
Testing Data
Feature extraction using MFCC
Feature extraction using LPCC
Testing with MFCC only
Testing with LPCC only

Fig. 1: Block diagram of the proposed method.

2. DATABASE FOR THE STUDY

Experiments are carried out on an our own created database of 30 speakers who can speak the three different languages. The database includes 17-male and 13-female speakers. The voice recording was done in the Engineering college laboratory. The speakers were undergraduate students and faculties 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 were 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 Beetel Head phone-250 with a frequency range 20-20 kHz. The speech files are stored in **.**wav format.

3. FEATURE EXTRACTION AND MODELING

Feature extraction is the most important part of speaker recognition. The purpose of feature extraction stage is to extract the speaker-specific information in the form of feature vectors at reduced data rate [2]. MFCC and LPCC based features have proven to be effective for Speaker recognition [8]. In this work, features are extracted using MFCC and LPCC techniques. Speech recordings were sampled at the rate of 8 kHz. Frame duration of 20 msec and a 10 msec for overlapping durations are considered. After framing, windowing (Hamming) method is carried out to minimize the spectral distortion. Fourier Transform (FFT) is then applied on the windowed frame signal to obtain the magnitude frequency response. The resulting spectrum is passed through a set of triangular band pass filters. We have considered 35 filters. In order to get the cepstral coefficients, Discrete cosine transform (DCT) is applied to the output of the mel filters. In this work, first 13 coefficients are considered as feature vectors. Since the 0^{th} coefficient can be regarded as a collection of average energies of each frequency bands, it is unreliable [20].

The LPCC can be easily obtained by Durbin's recursive procedure without computing the Discrete fourier transform (DFT) and the inverse DFT (IDFT), which are computationally complex and time consuming processes [21]. The vocal tract system is characterized by maximum of five resonances in the 0-4 kHz range, therefore, an LP order in the range 8-14 seems to be suitable for a speech signal sampled at 8 kHz [22]. The number of peaks of an all-pole system is determined by order of the LP analysis. In this study, speech is pre-emphasized (factor 0.97) to eliminate the radiation effect at the lips and then hamming windowed. 10 auto-correlation prediction coefficients were computed using the popular Levinson-Durbin algorithm and transformed into 13 cepstral coefficients.

VQ is a process of mapping vectors from large vector space to finite number of regions in the space. Most of the computation time in VQ-based speaker identification consists of distance computations between the unknown speaker's feature vectors and the models of the speakers enrolled in the system database [23]. In this work, the Linde-Buzo-Gray (LBG)-VQ technique is used with a splitting parameter (ε) of 0.05. The initial codebook is obtained by the splitting method and an initial code vector is set as the mean of the entire training data. This code vector is then split into two and the algorithm runs with these two codebooks. Later, these two code-books are split into four code-books and the iterative algorithm is repeated until the desired codebook size is achieved. We have generated different codebooks of sizes 64, 128 and 256.

The following steps are used in speaker identification process:

a) Choose the training data.
b) Extract the features using MFCC
c) Extract the features using LPCC
d) Combine MFCC and LPCC features
e) Generate the speaker model using VQ
f) Choose the testing data
g) Extract the features using MFCC separately
h) Extract the features using LPCC separately
i) Compare test features with speaker model
j) Use the Decision logic to find out the winner.

Table 1. : Different feature extraction techniques used for training and testing.

Expt.No	Training	Testing
1	MFCC	MFCC
2	LPCC	LPCC
3	MFCC + LPCC	MFCC
4	MFCC + LPCC	LPCC

4. MONOLINGUAL SPEAKER IDENTIFICATION SYSTEM

In monolingual speaker identification, training and testing languages are the same for a speaker [6] [18]. Since the data is collected in 3 different languages to study the robustness of the system, the experiments are conducted in three cases with the data of 15 seconds for training and testing. Each case includes four different experiments shown in Table 1.

Note that A/B indicates training with language A and testing with language B, for e.g., E/K indicates training with English language and testing with Kannada language. 3 cases, i.e.,

Case 1 : Training and testing with English language
Case 2 : Training and testing with Hindi language
Case 3 : Training and testing with Kannada language.

Case 1 : The results in Fig. 2(a) show that the speaker identification system yields good performance of 100% for codebook sizes of 128 and 256 when trained and tested with English language using MFCC+LPCC–VQ–LPCC method.

Case 2 : The results in Fig. 2(b) show that the speaker identification system yields good performance of 96.66 % for codebook sizes of 128 and 256 when trained and tested with Hindi language using MFCC+LPCC–VQ–LPCC method.

Case 3 : The results in Fig. 2(c) show that the speaker identification system yields good performance of 96.66% for codebook sizes of 128 and 256 when trained and tested with Kannada language using MFCC+LPCC–VQ–LPCC method.

5. CROSSLINGUAL SPEAKER IDENTIFICATION SYSTEM

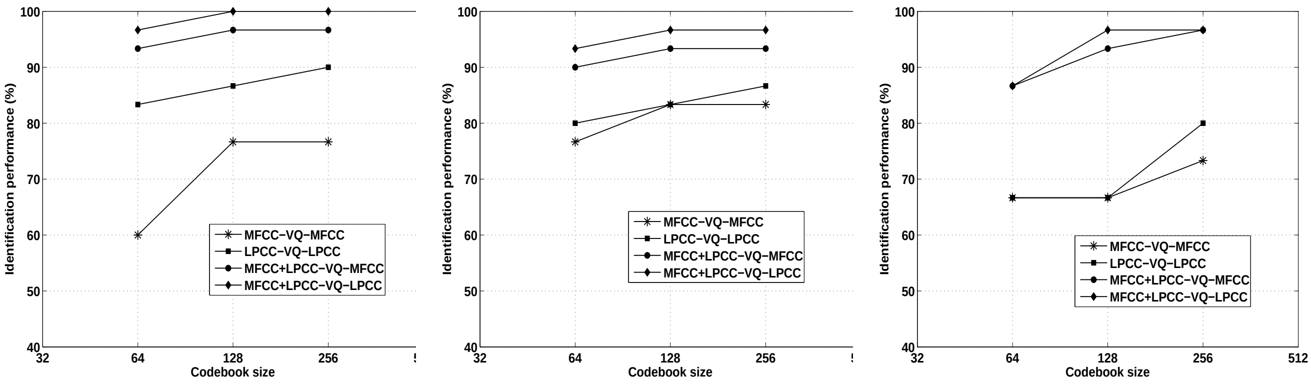
In crosslingual speaker identification (A/B), training is done in one language (say A) and testing is done in another language (say B) [6] [18]. Six different cases are considered in this context, i.e., E/H, E/K, H/E, H/K, K/E and K/H.

Case 1 (E/H) : The results in Fig. 3(a) show that the speaker identification system yields good performance of 83.33% for codebook size of 256 when training is done in English language and testing is done in Hindi language using the MFCC+LPCC–VQ–MFCC method.

Case 2 (E/K) : The results in Fig. 3(b) show that the speaker identification system yields good performance of 83.33% for codebook size of 256 when training is done in English language and testing is done in Kannada language using the MFCC+LPCC–VQ–LPCC method.

Case 3 (H/E) : The results in Fig. 3(c) show that the speaker identification system yields good performance of 90% for codebook size of 256 when training is done in Hindi language and testing is done in English language using MFCC+LPCC–VQ–LPCC method.

Case 4 (H/K) : The results in Fig. 4(a) show that the speaker identification system yields good performance of 73.33% for codebook size of 256 when training is done in Hindi language and

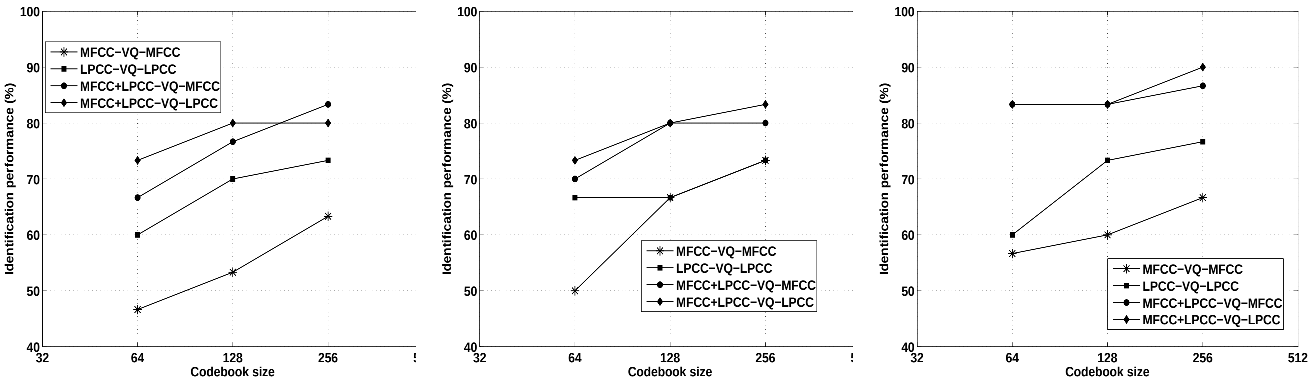


(a) E/E

(b) H/H

(c) K/K

Fig. 2: Performance of monolingual speaker identification.



(a) E/H

(b) E/K

(c) H/E

Fig. 3: Performance of crosslingual speaker identification.

testing is done in Kannada language using the MFCC+LPCC–VQ–MFCC method.

Case 5 (K/E) : The results in Fig. 4(b) show that the speaker identification system yields good performance of 90% for codebook size of 256 when training is done in Kannada language and testing is done in English language using the MFCC+LPCC–VQ–MFCC method.

Case 6 (K/H) : The results in Fig. 4(c) show that the speaker identification system yields good performance of 90% for codebook size of 256 when training is done in Kannada language and testing is done in Hindi language using the (MFCC+LPCC)–VQ–LPCC and (MFCC+LPCC)–VQ–MFCC methods.

6. MULTILINGUAL SPEAKER IDENTIFICATION SYSTEM

In multilingual speaker identification, some speakers in database are trained and tested in language A, some speakers are in language B and so on [6]. Arrangement of database for multilingual experiments is shown in Fig. 5. The speakers and languages are taken in random and there is no hard and fast rules for this arrangement.

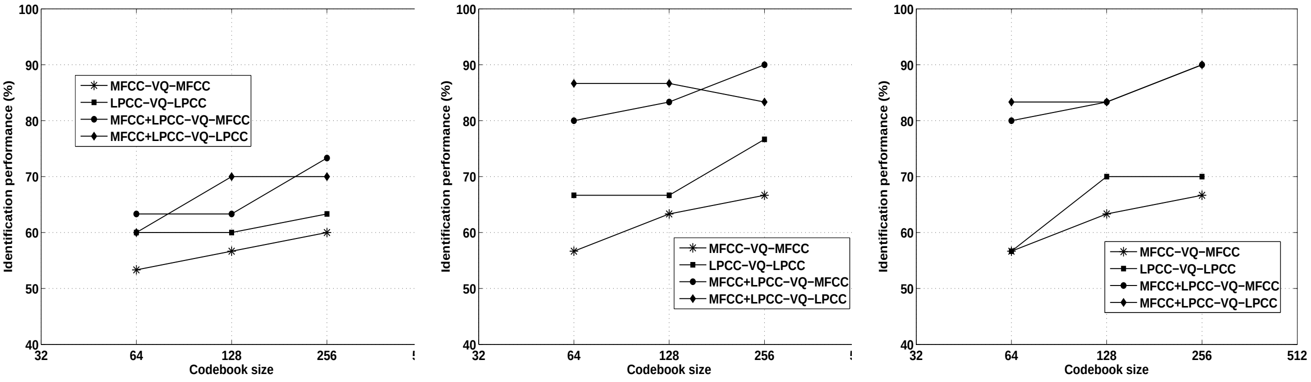
The results in the Fig. 6 show that the multilingual speaker identification system yields good performance of 100% for codebook sizes of 128 and 256 using the MFCC+LPCC–VQ–LPCC and MFCC+LPCC–VQ–MFCC methods. The combined (MFCC and LPCC) system identifies more number of speakers compared to all indi- vidual techniques and hence improvement in the performance. The improvement in performance is due to the different information provided by each feature.

The combined (MFCC and LPCC) system identifies some of the speakers which are not identified by MFCC and LPCC alone. For e.g., if the speaker 17 shown in the Table 2 is taken, speaker is not identified using MFCC or LPCC alone but is identified using the combined method. Note that if MFCC method is used, only 18 speakers could be identified, if LPCC method is used, then 20 speakers could be identified, but if the combined method is used, then 22 speakers could be identified.
Some of the observations can be made from the results are as follows:

(i) The proposed combination of features perform better in all the speaker identification experiments. This may be due to, different information provided by the MFCC and LPCC features (Fig. 7).

(ii) In comparison with the MFCC features, LPCC features perform better. This may be due to, LP-based features (LPCC) tract physiological characteristics of the vocal tract properties more effectively than the filter bank-based features [6].

(iii) In comparison with the monolingual speaker identification, crosslingual speaker identification performance decreases drasti-



(a) H/K

(b) K/E

(c) K/H

Fig. 4: Performance of crosslingual speaker identification.

Table 2. : Number of speakers identified by the MFCC-VQ-MFCC, LPCC-VQ-LPCC and MFCC+LPCC-VQ-LPCC systems for 30 speakers ($\surd$ identified ; x not identified).

Speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	No
MFCC	√	√	X	X	X	X	X	√	√	√	√	X	X	X	X	√	X	√	√	√	X	√	√	√	√	√	√	X	√	√	**18**
LPCC	√	√	√	√	X	X	X	√	√	√	√	√	√	X	X	√	X	√	√	X	X	√	√	√	√	√	√	X	√	X	**20**
Combined	√	√	√	√	X	X	√	√	√	√	√	X	√	X	X	√	√	√	√	X	X	√	√	√	√	√	√	X	√	√	**22**

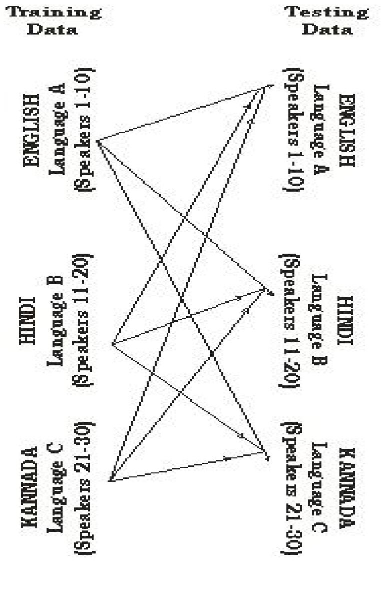

Training Data
Testing Data
ENGLISH Language A (Speakers 1-10)
HINDI Language B (Speakers 11-20)
KANNADA Language C (Speakers 21-30)
ENGLISH Language A (Speakers 1-10)
HINDI Language B (Speakers 11-20)
KANNADA Language C (Speakers 21-30)

Fig. 5: Multilingual Speaker Identification System Setup.

cally. This may be due to the variation in fluency and word stress when same speaker speaks different languages and also due to different phonetic and prosodic patterns of the languages [13].

(iv) The multilingual results are better than the crosslingual speaker identification experiments. This may be due to the better discrimination between the trained models and testing features (multiple languages) in multilingual scenario.

7. CONCLUSIONS

This work presented the task of mono, cross and multilingual speaker identification with the constraint of limited data condition using the combination of MFCC and LPCC features. 3 dimensional view confirm that the two feature extraction techniques namely MFCC and LPCC extract different information from the speech signal and hence can be fused so as to obtain a more robust speaker recognition system. The proposed combination of features perform better in all the speaker identification experiments. The results indicate that combination of features (MFCC+LPCC) can be used for improving the speaker identification performance in multilingual with the constraint of limited data. In order to study the robustness of the system needs to be verified with different languages (more than 3), different data sizes and more number of speakers.

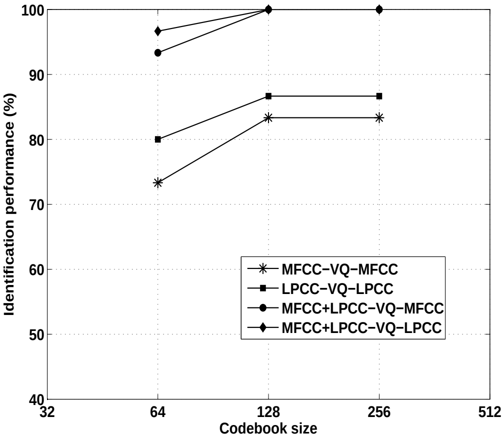


Fig. 6: Performance of multilingual speaker identification system.

8. ACKNOWLEDGMENT

This work is supported by Visvesvaraya Technological University, Belgaum–590018, Karnataka, India.

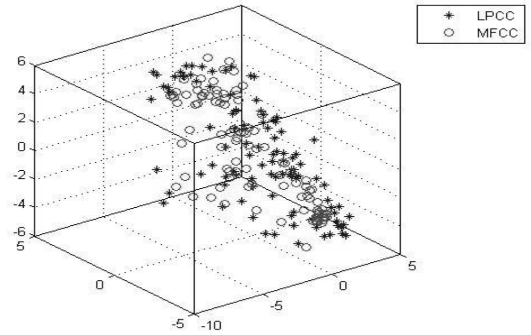


Fig. 7: The first 100 Features of a speaker for 15 secs speech data using MFCC and LPCC techniques in 3-Dimensional view.

9. REFERENCES

[1] B. S. Atal, “Automatic recognition of speakers from their voices,” Proc. IEEE, vol. 64(4), pp. 460475, Apr. 1976.

[2] H. S. Jayanna and S. R. Mahadeva Prasanna, “Analysis, Feature extraction, modeling and Testing techniques for Speaker Recognition,” IETE Technical Review, vol. 26, pp. 181190, 2009.

[3] J. P. Campbell, Jr.,“Speaker recognition: A tutorial,” Proc. IEEE, vol. 85(9), pp. 14371462, Sep. 1997.

[4] Bipul Pandey, Alok ranjan, Rajeev Kumar and Anupam Shukla, “Multilingual Speaker Recognition Using ANFIS,” in Proc. IEEE Int. Conf. Signal Processing Systems (ICSPS), (Dalian), pp. 714718, 2010.

[5] H. S. Jayanna, Limited data speaker recognition. PhD thesis, Indian Institute of Technology Guwahati, Dept. of Electronics and Communication Engg., Guwahati, India, Nov. 2009.

[6] P. H. Arjun, Speaker recognition in Indian languages: A feature based approach. PhD thesis, Indian Institute of Technology Kharagpur, Dept. of Electrical Engg., Kharagpur, India, Jul. 2005.

[7] Hemant A Patil, Sunayana Sitaram and Esha Sharma, “DA-IICT Crosslingual and Multilingual Corpora for Speaker Recognition,” Proc. IEEE, Advances in Pattern Recognition, (Kolkata), pp. 187190, 2009.

[8] Rama Murty and Yegnanarayana, “Combining Evidence from residual phase and MFCC features for Speaker Recognmition,” IEEE Signal Processing Letters., vol. 13(1), pp. 5255, Jan. 2006.

[9] Zhi-Yi LI, Liang HE,Wei-Qiang ZHANG and Jia LIU, “Multi-Feature Combination for Speaker Recognition,” Proc. IEEE, Chinese Spoken Language processing, (Tainan), pp. 318321, Dec. 2010.

[10] Xia Wang, Yang Cao, Feng Ding and Yuezhong Tang, “An embedded Multilingual speech recognition system for Mandarin, Cantonese, and English,” Proc. IEEE, Chinese Spoken Language processing, (Beijing, China), pp. 758764, Oct. 2003.

[11] Ulrike Halsband, “Bilingual and multilingual language processing,” Elseviers Journal of Physiology, (Paris), pp. 355369, 2006.

[12] Rajesh Ranjan, Sanjay Kumar Singh, Anupam Shukla and Ritu Tiwari, “Text-Dependent Multilingual Speaker Identification for Indian Languages using Artificial Neural Network,” Proc. Third International Conference on Emerging Trends in Engg. and Tech., (Goa), pp. 632 635, 2010.

[13] Geoffrey Durou, “Multilingual text-independent speaker identification,” Proc. MIST99 Workshop, (Leusden, Netherlands), pp. 115118, 1999.

[14] Prateek Agrawal, Anupam Shukla and Ritu Tiwari, “Multilingual speaker recognition using Artificial Neural network,” Advances in Intelligent and Soft Computing, vol. 116, pp. 19, 2009.

[15] Danoush Hosseinzadeh and Sridhar Krishnan, “Combining Vocal Source and MFCC Features for Enhanced Speaker Recognition Performance Using GMMs,” Proc. IEEE, MMSP-2007, (Crete), pp. 365368, Oct. 2007.

[16] Martine Adda Decker, “Towards the Multilingual interoperability concepts in automatic speech recognition,” Elseviers Speech communications, vol. 35, pp. 520, 2001.

[17] Olli Viikki, Imre Kiss and Jilei Tian, “speaker- and languageindependent speech recognition in Mobile communication systems,” Proc. IEEE, ICASSP01, vol. 1, (Salt Lake City, UT), pp. 58, 2001.

[18] Nagaraja B. G. and H. S. Jayanna, “Mono and Cross Lingual Speaker Identification with the constraint of Limited data,” Proc. IEEE, PRIME-2012, (Salem), pp. 439443, 2012.

[19] H. S. Jayanna and S. R. M. Prasanna, “Variable segmental analysis based speaker recognition in limited data conditions,” IEEE-Int. Conf. Signal, Image Process., vol. 2, (Karnataka, India), Dec. 2006.

[20] Picone J.W., “Signal modeling techniques in speech recognition,” Proc. IEEE, vol. 81(9), pp. 12151247, 1993.

[21] A comparison of speech recognition ability between LPCC and MFCC. Proc. of the National Systems Conference, NSC 95-2221-E-451-014.

[22] S. R. M. Prasanna, C. S. Gupta and B. Yegnanarayana, “Extraction of speaker-specific excitation information from linear prediction residual of speech,” Speech Communication, vol. 48, pp. 12431261, 2006.

[23] G. Senthil Raja, Feature analysis and compensation for speaker recognition under stressed condition. PhD thesis, Indian Institute of Technology Guwahati, Dept. of Electronics and Communication Engg., Guwahati, India, Jul. 2007.