Analysis of Language Identification Performance based on Gender and Hierarchial Grouping Approaches

B. Bhaskar, Dipanjan Nandi and K Sreenivasa Rao
School of Information Technology Indian Institute of Technology Kharagpur, India - 721302
bhass510@gmail.com, dipanjannandi@sit.iitkgp.ernet.in, ksrao@iitkgp.ac.in

Abstract

In this work, language identification (LID) performance is analyzed in gender independent, gender dependent and hierarchical grouping approaches. In hierarchical grouping approach, three language groups are developed based on the confusion patterns of languages obtained from gender independent approach. In the hierarchical grouping LID approach, the language group of a test utterance is recognized at first level, and then the particular language within the group is identified. In gender based LID system, the gender of a test utterance is identified first, and then it is evaluated with all the language models of corresponding gender to identify the language of test utterance. The vocal tract system information represented by mel-frequency cepstral coefficients (MFCCs) is used to capture the language-specific information in this work. The Gaussian mixture models (GMMs) are used to develop the language models. The Indian Institute of Technology Kharagpur - Multi Lingual Indian Language Speech Corpus (IITKGP-MLILSC) is used to carry out the LID study. The LID performance of gender independent system is 74.47%. The LID performance of hierarchical grouping approach is 72.35% which is slightly lower than the gender independent LID system. However, the LID accuracy of the gender based system is 76.58% which is relatively better than other two approaches.

***I*ndex Terms—** Gender independent LID, Gender dependent LID, Hierarchical grouping, MFCCs, IITKGP-MLILSC.

1 Introduction

Speech contains sequence of sound units. These sound units have the respective language specific constraints and are also influenced by speaking style. So, the human speech signal carries information about message, speaker, language and also emotion of the speaker. The goal of automatic language identification (LID) task is to determine the language from the uttered speech accurately. Each language has unique syntax and speaking style. The languages of India are divided into two major groups, the Indo-Aryan languages and the Dravidian languages. Almost all the Indian languages share the common set of phonemes. Therefore, it is difficult to develop an automatic language identification system for Indian languages accurately. Due to several practical applications like information retrieval from multilingual databases, speech to speech translation and voice activated systems, exploration on LID task by machine has drawn a great deal of interest and attention. Human speech production mechanism can be modeled as a time-varying vocal tract filter excited by time-varying excitation source. Thus, both vocal tract and excitation source information is reflected in the speech signal. In recent works, vocal tract information explored using block processing method, pitch synchronous approach (PSA) and glottal closure region (GCR) based approach (K. S. Rao, 2013). Prosody information also used for language identification task (V. R. Reddy, 2013). In this work, vocal tract information represented by mel-frequency cepstral coefficients (MFCC) (S. B. Davis, 1980) is used to capture the language-specific information. The LID study is carried out on IITKGP-MLILSC (S. Maity, 2012) database. The gaussian mixture models (GMMs) (D. A. Reynolds, 1995) are used to capture the distribution of language-specific information. In this study, language identification has been studied in gender independent, gender dependent and hierarchial grouping approaches. In hierarchial grouping approach, the language group of a test utterance is decided first, and then the language within that language group is identified. The spectral shape of human vocal tract system is different for male and female. Therefore, we have carried out the LID experiment based on gender based approach. In this approach, the gender of a test utterance is first recognized and it is then evaluated to all the language models of the corresponding gender to identify the language of the test utterance. The LID performance of the hierarchial based approach is slightly lower than the gender independent language identification approach, whereas, the gender dependent language identification study provides better LID performance.

The rest of the paper is organized as follows : Section 2 explains the vocal tract system features which have been used in this work. In Section 3, the development of language models using GMM is described. In Section 4, the detail description of language database has been given. In Section 5, the development procedures of three different LID systems and performances are analyzed. Summary and conclusion of the present work are discussed in Section 6.

2 Features for LID

The human speech production system consists of a vocal tract and a source for exciting the vocal tract resonator. During speech production, vocal tract system behaves like a time varying resonator or may be treated as a time varying filter. This time varying filter characterizes the variations in the vocal tract shape in the form of resonances and anti-resonances that occur in the speech spectrum. Parameterization techniques like, linear prediction cepstral coefficients (LPCCs) and mel-frequency cepstral coefficients (MFCCs) (S. B. Davis, 1980) are available for modeling vocal tract information. Since mel-filters are based on human perceptual nature, we have used MFCC feature for this LID study. The steps for calculating the MFCCs from the speech signal are discussed below.

(i) Pre-emphasis : It refers to a filtering technique that emphasizes the higher frequencies. Some voiced sounds have a steep roll-off in the high frequency region. So, to balance the speech spectrum of voiced sounds, high-frequency filtering is needed. The procedure for performing the pre-emphasis is shown in the equation 1.

$$H(z) = 1 - \alpha z^{-1} \tag{1}$$

where the value of α controls the slope of the filter and is usually between 0.9 to 1.0.

(ii) Windowing : The human speech signal is a quasi-stationary signal. The voiced sound units are quasi-periodic in nature, whereas, the unvoiced sound units are noise like signal. Therefore, the analysis of speech signal for any speech applications must always be carried out on short segments across which the speech signal is assumed to be stationary. Short-term spectral measurements are typically carried out over the range of 10-30 ms frame size and frame shift of half of the frame size. The blocked frames are Hamming windowed. Hamming window is used to reduce the edge effect while taking the discrete Fourier transform (DFT) on the signal.

(iii) Discrete Fourier Transform (DFT) : Each windowed frame is converted into magnitude spectrum by applying DFT using the equation 2.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi nk}{N}}, \quad 0 \leq k \leq N-1. \tag{2}$$

where $x(n)$ is the samples of the windowed speech signal. $X(k)$ is the magnitude spectrum of windowed speech signal and N is the number of points used to compute the DFT.

(iv) Mel-spectrum : The mel-spectrum is computed by passing the DFT spectrum through a set of band-pass triangular filters known as mel-filter bank. A mel is a unit of perceived speech frequency or a unit of tone. The mel scale is therefore a mapping between the physical frequency scale (Hz) and the perceived frequency scale (Mels). The approximation of mel from physical frequency can be expressed by the following equation (S. B. Davis, 1980)(J. R. Deller jr., 2000).

$$f_{mel} = 2595 \log(1 + \frac{f}{700}) \tag{3}$$

where f denotes the physical frequency and f_{mel} denotes the perceived mel-frequency. The mel-spectrum values or mel-frequency coefficients of the magnitude spectrum $X(k)$ is computed by multiplying the magnitude spectrum by each of the triangular mel-weighting filters.

$$S(m) = \sum_{k=0}^{N-1} |X(k)|^2 H_m(k), \quad 0 \leq m \leq M-1. \tag{4}$$

where $S(m)$ is the mel-frequency coefficients and M is total number of triangular mel-weighting filters.

(v) Inverse Discrete Cosine Transform (IDCT) : The $\log$ operation is performed on the mel-frequency coefficients. The IDCT is then applied to obtain the cepstral coefficients. This yields a signal in the cepstral domain. MFCC is computed as follows :

$$c(n) = \sum_{m=0}^{M-1} \log(S(m)) \cos(\frac{\pi n(m-0.5)}{M}), \quad n = 0, 1, 2, ..., C-1 \tag{5}$$

where $c(n)$ are the cepstral coefficients and C is the number of MFCCs. The zeroth coefficient represents the average log-energy of the input signal.

3 Development of Language Models

In this LID study, the acoustic features are modeled by Gaussian probability density functions (PDFs), described by the mean vector and the covariance matrix. However unimodel PDF with only one mean and covariance are unsuitable to model all variations of a single event in speech signals. Therefore, a mixture of single densities is used to model the complex structure of the density probability. For a D-dimensional feature vector denoted as x_t, the mixture density for language s is defined as weighted sum of M component Gaussian densities as given by the following equation (D. A. Reynolds, 1995)

$$P(x_t|s) = \sum_{i=1}^{M} w_i P_i(x_t) \tag{6}$$

where w_i are the weights and $P_i(x_t)$ are the component densities. Each component density is a D-variate Gaussian function of the form

$$P_i(x_t) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}[(x_t-\mu_i)'\Sigma_i^{-1}(x_t-\mu_i)]} \tag{7}$$

where μ_i is a mean vector and Σ_i covariance matrix for i^th component. The mixture weights have to satisfy the constraint (D. A. Reynolds, 1995)

$$\sum_{i=1}^{M} w_i = 1 \tag{8}$$

The complete Gaussian mixture density is parameterized by the mean vector, the covariance matrix and the mixture weight from all component densities. These parameters are collectively represented by

$$s = \{w_i, \mu_i, \Sigma_i\}; \quad i = 1, 2, ..., M \tag{9}$$

To determine the model parameters of GMM of a particular language, the GMM has to be trained. In the training process, the maximum likelihood (ML) procedure is adopted to estimate model parameters. The main objective of the ML estimation is to derive the optimum model parameters that can maximize the likelihood of GMM. The likelihood value is, however, a higher order nonlinear function in the model parameters and therefore, direct maximization is not possible. Instead, maximization is done through iterative procedures. The most popular maximization technique is the iterative expectation maximization (EM) algorithm (A. P. Dempster, 1977). In our work EM algorithm has been exploited to obtain the optimal model parameters. The EM algorithm begins with an initial model s and tends to estimate a new model such that the likelihood of the model increases with each iteration. This new model is considered to be an initial model in the next iteration and the entire process is repeated until a certain convergence threshold is obtained or a certain predetermined number of iterations have been made. The performance of EM algorithm depends on the initialization. In our work, we have exploited K-means clustering algorithm (Y. Linde, 1980) for initializing the GMM model parameters. In each iteration the posterior probabilities for the i^{th} mixture is computed which is given as follows (D. A. Reynolds, 1995) :

$$Pr(i|x_t) = \frac{w_i P_i(x_t)}{\sum_{j=1}^{M} w_j P_j(x_t)} \tag{10}$$

The model parameters are updated according to the following expressions (D. A. Reynolds, 1995) :

The updated mixture *weight* is

$$\overline{w_i} = \frac{\sum_{t=1}^{T} Pr(i|x_t)}{T} \qquad (11)$$

The updated *mean* vector is

$$\overline{\mu_i} = \frac{\sum_{t=1}^{T} Pr(i|x_t)x_t}{\sum_{t=1}^{T} Pr(i|x_t)} \qquad (12)$$

The updated *covariance matrix* is

$$\overline{\sigma_i^2} = \frac{\sum_{t=1}^{T} Pr(i|x_t)|x_t - \overline{\mu_i}|^2}{\sum_{t=1}^{T} Pr(i|x_t)} \qquad (13)$$

where, T denotes the total number of feature vectors in the training set of a language. In the estimation of the model parameters, it is possible to choose, either full covariance matrices or diagonal covariance matrices. It is more common to use diagonal covariance matrices for GMM, since linear combination of diagonal covariance Gaussians has the same model capability with full matrices (Q. Hong, 2005). Another reason is that speech utterances are usually parameterized with cepstral features. Cepstral features are more compactable, discriminative, and most important, they are nearly uncorrelated, which allows diagonal covariance to be used by the GMMs (D. A. Reynolds, 1995). In our work iterative process has been carried out 50 times at which point the model is presumed to be converged to a local maximum.

4 Language Database

In this work, LID study has been carried out on Indian Institute of Technology Kharagpur - Multi Lingual Indian Language Speech Corpus (IITKGP-MLILSC) (S. Maity, 2012). This database contains 27 Indian regional languages. Sixteen languages are collected from news bulletins of broadcasted radio channels and the remaining are recorded from broadcasted TV talk shows, live shows, interviews and news bulletins. Each language comprises of minimum 1 hour of speech data from at least 10 speakers that includes both male and female speakers. For each speaker 5-10 minutes data is collected at the sampling rate of 16 kHz with 16 bits per sample. The speech signal is down-sampled to 8 kHz for our LID study. We have used average 45 minutes of data from each language for developing the language models. 28 test utterances each of 10 sec duration have been used form each language to evaluate the language models. The broadcasted television channels are accessed using VentiTV software and the Pixelview TV tuner card. Audacity software is used for recording the speech data from TV channels. The language data of broadcasted Radio channels are collected from the archives of Prasar Bharati, All India Radio (AIR) website.

5 Development and Performance Evaluation of Proposed Language Identification Systems

In this work, we have developed three different LID systems : (a) gender independent LID system, (b) hierarchial group based LID system and (c) gender dependent LID system. In this Section 5.1, 5.2 and 5.3 the details of gender independent, hierarchial group based and gender dependent LID systems and their performance evaluation are described.

Table 1: Performance of Gender Independent LID System

Languages	Average Recognition Performances (%)			
	Rank 1	Rank 2	Rank 3	Rank 4
Arunachali	100	100	100	100
Assamese	100	100	100	100
Bengali	46.42	67.85	85.71	96.42
Bhojpuri	0	0	0	3.57
Chhatisgarhi	100	100	100	100
Dogri	53.57	89.28	96.42	100
Gojri	92.85	100	100	100
Gujarati	89.28	100	100	100
Hindi	100	100	100	100
Indian English	50	50	50	50
Kannada	100	100	100	100
Kashmiri	100	100	100	100
Konkani	85.71	100	100	100
Malayalam	39.28	82.14	85.71	100
Manipuri	96.42	96.42	96.42	96.42
Marathi	10.71	28.57	42.85	75
Mizo	96.42	96.42	100	100
Nagamese	100	100	100	100
Nepali	60.71	78.57	89.28	96.42
Oriya	25	28.57	39.28	50
Punjabi	96.42	100	100	100
Rajasthani	100	100	100	100
Sanskrit	57.14	92.85	100	100
Sindhi	39.28	50	64.28	89.28
Tamil	100	100	100	100
Telugu	100	100	100	100
Urdu	71.42	100	100	100
Average	**74.47**	**83.73**	**87.03**	**91**

Table 2: Confusion Matrix of Gender Independent LID System (Rows indicate the classification of test utterances of the corresponding language and the columns indicate the test utterances falling under the corresponding language. The columns follow the same order of languages as rows.)

Languages	Average Recognition Performances (%)																									
Arunachali	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Assamese	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bengali	0	0	46	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	50	0	0	0	0	0	0	0	0
Bhojpuri	7	21	0	0	0	0	0	0	0	0	0	0	0	3	0	67	0	0	0	0	0	0	0	0	0	0	0
Chattisgarhi	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dogri	46	0	0	0	0	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gojri	0	0	0	0	0	0	92	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gujarati	10	0	0	0	0	0	0	89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hindi	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Indian English	3	39	0	0	0	0	0	0	0	50	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
Kannada	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kashmiri	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Konkani	0	0	0	0	0	0	14	0	0	0	0	0	85	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Malayalam	0	7	0	0	0	0	0	0	0	0	0	7	0	39	0	25	0	0	0	21	0	0	0	0	0	0	0
Manipuri	0	0	0	0	0	0	0	3	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0
Marathi	0	0	39	0	0	0	0	0	0	0	0	0	0	28	0	10	0	0	0	21	0	0	0	0	0	0	0
Mizo	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0
Nagamese	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
Nepali	0	3	0	0	0	0	0	3	0	0	0	0	0	21	0	0	0	7	60	0	0	0	0	0	3	0	0
Oriya	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	10	0	0	14	25	0	0	0	0	0	0	0
Punjabi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	3
Rajasthani	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
Sanskrit	35	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57	0	0	0	0
Sindhi	0	0	0	0	0	57	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0
Tamil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
Telugu	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
Urdu	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	3	0	0	71

5.1 Gender Independent Language Identification System

The gender independent LID study has been carried out on 27 Indian languages. The language models are developed by considering both male and female speakers of a particular language. The Gaussian mixture models (GMMs) (D. A. Reynolds, 1995) are used to build the language models. Different Gaussian mixtures (4, 8, 16, 32, 64, 128 and 256) have been explored for modeling the language-specific knowledge. For evaluating the performance of developed LID systems, two speakers from each language who have not participated during training phase are considered. 28 test utterances from each language, each of 10 sec duration have been considered for evaluation of the language models. 14 test utterances are collected from a male speaker and remaining 14 are from a female speaker for evaluation. The performance for optimum Gaussian mixture of gender independent LID system is given in Table 1. The average LID performance of 27 languages is 74.47%.

In Table 1, the individual language performances for gender independent LID systems are shown. The individual language performances are examined at rank - 1, rank - 2, rank - 3 and rank - 4 positions. A test utterance is evaluated against all 27 language models and each model gives probability scores. Rank - 1 position indicates the position of the maximum probability score among all 27 scores. Rank -2 performance is calculated by considering top two probability scores. Similarly, rank - 3 and rank - 4 performances are calculated by considering the top three and top four probability scores among 27 scores supplied by the language models respectively. Second column of Table 1 represents the LID performance at rank - 1 position. By examining the performance of individual language, it can be observed that the LID accuracy of 10 languages are 100% and the remaining languages are misclassified with other languages. We have developed the hierarchial group based LID system by analyzing the confusion patterns of the languages in gender independent LID system at rank - 1 position. In Section 5.2, the LID performance is analyzed based on grouping of different languages. The LID performances for Gojri, Gujarati, Konkani, Manipuri, Mizo and Punjabi languages belongs to the range of (80-100)%. The LID accuracy of Dogri, Indian English, Nepali, Sanskrit and

Urdu languages belongs to the range of (50-80)%. The LID performances of remaining languages are not significant. From Table 1, it is also observed that, the average identification performance has been increased from 74.47% to 91% by considering top 4 ranks among the 27 languages. The LID performances at rank 1 position for Bengali, Bhojpuri, Dogri, Malayalam, Marathi, Nepali, Oriya, Sindhi and Urdu languages are 46.42%, 0%, 53.57%, 39.28%, 10.71%, 60.71%, 25%, 39.28% and 71.42%. The LID performances are improved significantly for these languages by considering rank 4 position.

Table 3: Performance of Gender Dependent LID System

Languages	Average Recognition Performances (%)
Arunachali	100
Assamese	96.42
Bengali	50
Bhojpuri	0
Chhatisgarhi	100
Dogri	100
Gojri	67.85
Gujarati	67.85
Hindi	96.42
Indian English	50
Kannada	100
Kashmiri	92.85
Konkani	100
Malayalam	42.85
Manipuri	96.42
Marathi	14.81
Mizo	100
Nagamese	100
Nepali	53.57
Oriya	14.28
Punjabi	100
Rajasthani	100
Sanskrit	85.71
Sindhi	39.28
Tamil	100
Telugu	100
Urdu	100
Average Performance	**76.58**

5.2 Hierarchial Group Based Language Identification System

The confusion matrix of gender independent LID system using 27 languages is shown in Table 2. Rows indicate the classification of test utterances of the corresponding language and the columns indicate the test utterances falling under the corresponding language. The diagonal elements represent the correctly identified performances for corresponding languages. The columns follow the same order of languages as rows. Three language groups are made by analyzing the confusion matrix of gender independent system shown in Table2. In first group we have kept the 5 languages which does not confuse with other languages and provides 100% LID performance. The languages belongs to the first group are, Chhatisgarhi, Hindi, Kannada, Tamil and Telugu. The other two groups contains the languages which are confused with other languages within that group mostly. The second group consists of Arunachali, Dogri, Gojri, Konkani, Manipuri, Punjabi, Rajasthani, Sanskrit, Sindhi and Urdu. The third group comprises of Assamese, Bengali, Bhojpuri, Indian English, Kashmiri, Malayalam, Marathi, Mizo, Nagamese, Nepali and Oriya. The average group level performance is 96.82% and the average LID performance of 27 languages are 72.35% which is slightly lower than the gender independent LID system. The LID performances for Assamese, Hindi, Kannada, Kashmiri, Rajasthani, Tamil and Telugu languages are 100%. The LID performances for Arunachali, Gojri, Gujarati, Konkani, Manipuri, Mizo, Nagamese, Punjabi and Urdu languages belongs to the range of (80-100)%. The LID accuracy of Bengali, Chhatisgarhi, Dogri, Indian English, Sanskrit and Sindhi languages belongs to the range of (40-80)%. The performances of the remaining languages are negligible.

5.3 Gender Dependent Language Identification System

Two language models are developed for each language based on gender. The gender of the test utterances are identified first and then evaluated to all 27 language models of the corresponding gender. The LID performance of individual languages are given in the Table 4. The average LID accuracy of 27 languages is 76.58%. The LID performance is relatively better compares to gender independent and hierarchial group based LID systems. The LID performances for Arunachali, Chhatisgarhi, Dogri, Kannada, Konkani, Mizo, Nagamese, Punjabi, Rajasthani, Tamil, Telugu and Urdu languages are 100%. The identification performances for Assamese, Gojri, Gujarati, Hindi, Kashmiri, Manipuri, Nepali and Sanskrit languages are belongs to the range of (60-100)%. The identification accuracy of Bengali, Indian English, Malayalam and Nepali languages belongs to

Table 4: Performance of Hierarchial Group Based LID System

Languages	Average Recognition Performances (%)
Arunachali	96.42
Assamese	100
Bengali	42.85
Bhojpuri	0
Chhatisgarhi	57.14
Dogri	50
Gojri	92.85
Gujarati	82.14
Hindi	100
Indian English	50
Kannada	100
Kashmiri	100
Konkani	82.14
Malayalam	25
Manipuri	96.42
Marathi	14.28
Mizo	96.42
Nagamese	92.85
Nepali	21.42
Oriya	32.14
Punjabi	96.42
Rajasthani	100
Sanskrit	53.57
Sindhi	50
Tamil	100
Telugu	100
Urdu	85.71
Average Performance	**72.35**

the range of (40-60)%. The remaining languages does not provide significant LID performances.

6 Summary and Conclusion

In this work, we have analyzed the language identification performance based on gender independent, hierarchial grouping and gender dependent approaches. Vocal tract features has been used to capture the language-specific information in this work. In gender independent LID system, we have examined the individual language performance as-well-as average LID performance by considering top 4 ranks of 27 Indian languages. The average LID performance of gender independent system at rank - 1 and rank - 4 position are 74.47% and 91% respectively. The hierarchial group based LID system has been developed based on the confusion patterns of gender independent LID system. The identification performance of hierarchial group based system and gender dependent system are 72.35% and 76.58% respectively. The LID performance of gender dependent LID system is better among all three approaches. The reason is that, the dynamics of vocal tract system is significantly different for male and female. The hierarchial group based LID system has been developed is slightly lower than the gender independent LID system. The reason is that, the identification accuracy at the group level is not 100%. Therefore, the overall LID accuracy has not improved for hierarchial group based LID system. Vocal tract features are exploited only for this study. In future, excitation source information along with vocal tract information can be explored to develop language identification system.

References

K. S. Rao, S. Maity, and V. R. Reddy. 2013. *Pitch synchronous and glottal closure based speech analysis for language recognition*. International Journal of Speech Technology (Springer), vol. DOI: 10.1007/s10772-013-9193-5.

V. R. Reddy, S. Maity, and K. S. Rao. 2013. *Identification of indian languages using multi-level spectral and prosodic features*. International Journal of Speech Technology (Springer), vol. DOI: 10.1007/s10772-013-9198-0.

S. B. Davis and P. Mermelstein. 1980. *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*. IEEE Trans. Acoust. Speech and Signal Processing, vol. 28, no. 28, pp. 357–366.

S. Maity, A. K. Vuppala, K. S. Rao, and D. Nandi. 2012. *IITKGP-MLILSC speech database for language identification*. National Conference on Communication.

D. A. Reynolds and R. C. Rose. 1995. *Robust text -independent speaker identification using gaussian mixture speaker models*. IEEE Trans.Speech and Audio Proc., vol. 3, no. 1, pp. 72–83.

J. R. Deller jr., John H. L. Hansen and J G Proakis. 2000. *Discrete-Time Processing of Speech Signal*. 2nd ed. New York: IEEE Press.

A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society, Series B (Methodological), vol. 39, no. 1, pp. 1–38.

Y. Linde, A. Buzo and R. M. Gray. 1980. *An algorithm for vector quantizer design*. IEEE Trans. on Communication, vol. 28, pp. 84–95.

Q. Hong and S. Kwong. 2005. *A discriminative training approach for text-independent speaker recognition*. Signal Processing, vol. 85, no. 7, pp. 1449 – 1463.