Analysis of Different Feature for Language Identification

Snehal V. Gite
P.G. Student.
Dept. of Computer Engineering
Late G.N. Sapkal
College of Engineering Nasik,
India

J. V. Shinde
Assistant Professor
Dept. of Computer Engineering
Late G.N. Sapkal
College of Engineering Nasik,
India

ABSTRACT

Language Identification is the task of identifying language spoken from unknown user. The main objective is to achieve accurate results in shortest speech segments by using automatic Language Identification system. It works on language classification that involves new language rapid learning identities and reduce the computational complexity. MFCC, GFCC, PLP and the combination of these feature are consider in language identification system. The proposed approach that transforms the spoken words to a represent low dimensional i-vector, on which classification techniques are applied. Feature extraction is done on input audio, Universal background model and i-vector extraction are used in proposed system in order to meet the challenges involved in rapidly making reliable decisions about the spoken language such as Marathi, Hindi and English. For the relevant languages under the different acoustic condition are used to capture robust feature extraction scheme.

Keywords

Language Identification, Feature Extraction, Universal background model.

1. INTRODUCTION

Automatic language identification is the process of identifying the language i.e identify speech spoken by an unknown speaker [1] [2]. For speech recognition people are the most precise language ID frameworks. Inside of seconds of listening to discourse, individuals can figure out if it is a language they know or not. In the event that it is a language with they are not well known, they frequently can make subjective judgments as its closeness to a language they know, Language ID (Cover) is the procedure of distinguishing the talked language from the recording. When the language is new to human the subjective decision are made but it turn out to be less reliable. So aim of this paper to contribute in this area in which decision about the spoken language should be made rapidly in short duration words and precise inside of a few moments.

For the task of text-independent speaker verification, likelihood ratio detection using Maximum A-Posteriori (MAP) used Gaussian Mixture Models (GMMs) and Universal Background model (UBM) has become the standard approach. While this approach provides very good performance, a continuing challenge for robust speaker verification is dealing with channel or session variability [3]. JFA involves UBM to reduce the variability from non-language related effect. JAF is mainly used for the speaker verification. So the aim of robust language identification on short utterances to identify language with accurate accuracy in short duration, by adopting extracted I-vector model and universal background model for training and testing phase.

This work contributes challenging problem of language identification system. Main objective is to achieving a high accuracy on a small amount of data.

Paper formulates the problem on how to minimize the computational complexity of identified language. It also proposed LID for short utterances in for short duration to improve the accuracy and performance.

2. LITERATURE SURVEY

Language ID utilizing phoneme acknowledgment and phonotactic demonstrating takes after by n-gram language models and uses Language ID utilizing phoneme acknowledgment and phonotactic demonstrating takes after by n-gram language models and uses PRLM. It present gender-dependent acoustic model. This method is utilized for enhancing speech recognition performance. Yet, because of gender dependent accuracy precision is low so this framework can enhance exactness for the gender dependent.[4]

Language Identification (LID) taking into account language-dependent telephone acknowledgment utilizes various elements and their mixes that are extricated by language-dependent recognizers were assessment depends on the same database. Two techniques are utilized that are [5]

- Forward and backward bigram based language models
- Context-dependent duration models.

Both the methods are used for language identification, backward bigram is used to capture backward phonetic constrains only.

SVM based speaker confirmation utilizing GMM Model, Gaussian mixture models with universal backgrounds (UBMs) have turned into the standard strategy for speaker recognition. A speaker model is model by MAP adjustment of the method for the UBM. A GMM supervector is developed by the method for received mixture components. A late research is that element investigation of this GMM supervector is a viable strategy for variability remuneration. System build a support vector machine utilizing strategy called as GMM supervector[6].

Acoustic, phonetic and discriminative approaches deal with programmed language identification that presented 3 techniques GMM, phone recognition and support vector machine arrangement yet rectify exactness is not achieved [7].

Total variability model utilizes i-vector approach based on Joint Factor Analysis for speaker verification. JFA model based on speaker and channel components comprises of two particular spaces: the speaker space element characterized by the eigenvoice matrix V and the channel space element represent by the eigenchannel matrix U .Only single space is utilized rather than two, which refers to as the total variability space. The total variability network contains the eigenvectors that have the biggest eigenvalues of the total variability covariance matrix. Given an utterances , the new speaker-and channel-subordinate GMM supervector characterized as takes after:

$$M = m + Tw \qquad (1)$$

Joint factor analysis versus Eigen channels in speaker recognition this procedure introduced two ways to deal with the issue of session variability in Gaussian mixture model (GMM)- based speaker verification, Eigechannels, and joint factor examination that component investigation was significantly more viable than Eigenchannel modeling.[8] In proposed system we use JFA methods for accuracy in speaker verification task.[8]

For speaker confirmation front-end factor analysis speak to another speaker check framework where further analysis is utilized to define new low-dimensional space that models both speaker and channel variability. [9]

I-vectors in the connection of phonetically-constrained short utterance for speaker verification future extent of this method is to exploring the effect of phonetic data on i-vector standardization by considering the relationship between's the current speaker segregation scoring and distinction between expressions phonetic separation for short duration[10]

Language recognition in i-vectors space The idea of alleged i-Vectors, where every utterance is spoken to by fixed length low-dimensional element vector, novel methodology for language recognition that model gives magnificent execution over all conditions future extension is attempt to acquire i-Vectors from the utterances and the relating adequate insights in a more straightforward manner.

The methods based on I vector model for speaker recognition presents a study for how the current selection of factor analysis techniques that perform when utterance lengths are significantly reduced. Problems of short utterance with factor analysis approaches will be investigated in future.

3. PROPOSED WORK

For the language identification number of methods are used but some drawbacks are there .So for the accurate identification in short duration is the main aim of this paper. Fig 1. Shows proposed system architecture. In proposed system first input audio is taken then preprocessing is done i.e. speech enhancement, audio should capture the acoustic properties. In language identification as the change in background noise can change in acoustic condition. Voice Activity Detection (VAD) is used to prevent non-speech audio segments from interfering with the classification decision, a speech enhancement method are used for noise deformation and a robust feature extraction module. To reduce the sensitivity of the features to the acoustic variability normalization step is done. Simplified i-vector model for the LID is used on both training and testing phase by doing feature extraction. For the high accuracy and performance on a small amount of speech data is the main aim of this paper. In GMM-based Language identification, from the acoustical representation statistical probabilities are derived that the spoken words will be accumulated over time and that will be propagated until the final classification stage. Performance also increases when more statistics can be accumulated, these tend to be system more robust on short utterances where they do not based on rule-based approaches applied on phonetic transcription. For the accurate language identification i-vector frame work is used in proposed system[13][14]. To reduce the computational complexity RLID adopts the simplified i-vector framework and UBM fused total variability model. These simplified work of i-vector system slightly reduces the complexity of the conventional i-vector baseline[15].

3.1 Feature Extraction

Proposed system extract the different acoustic feature that Mel Frequency cepstral coefficients (MFCC), Perceptual linear prediction (PLP), Gammatone Frequency Cepstral Coefficients (GFCC), and the combination of these feature.

3.2 Simplified I-vector extraction model

For the multifaceted nature reduction in proposed framework utilize simplified i-vector model. This simplified and supervised i-vector model is applied in the task of robust and efficient speaker verification. In that first linking the mean supervector and the i-vector variable factor loading matrix with the label vector and the direct classifier framework, the i-vectors then extended to label-regularized supervised i-vectors. These supervised i-vectors are advanced to remake the mean supervectors and minimize the mean squared error between the first and the recreated name vectors, such that they turn out to be more discriminative. Second, figure examination (FA) can be performed on the pre-standardized focused GMM first request insights supervector is utilized to guarantee that the Gaussian measurements sub-vector of each Gaussian segment is dealt with similarly in the FA, which decreases the computational cost essentially.In the Simplified modeling adopted in both training and testing phase.it requires prenormalization step that reduce the complexity in i-vector extraction. By applying re-weighting schema simplified model can be given as,

$$F_j^c = \sqrt{\frac{B_j^c}{b_j^c}}\, F_j^c \qquad (2)$$

Where B_j^c is diagonal matrix which is composed of c diagonal blocks. F_j^c is supervector.

3.3 Methodology

In this system can be work in phases i.e training and testing phase

Methodology for Training phase

1. 1 .Input Signal.
2. Convert to Vocal feature.
3. Apply feature extraction methods.
4. Apply UBM model for training.

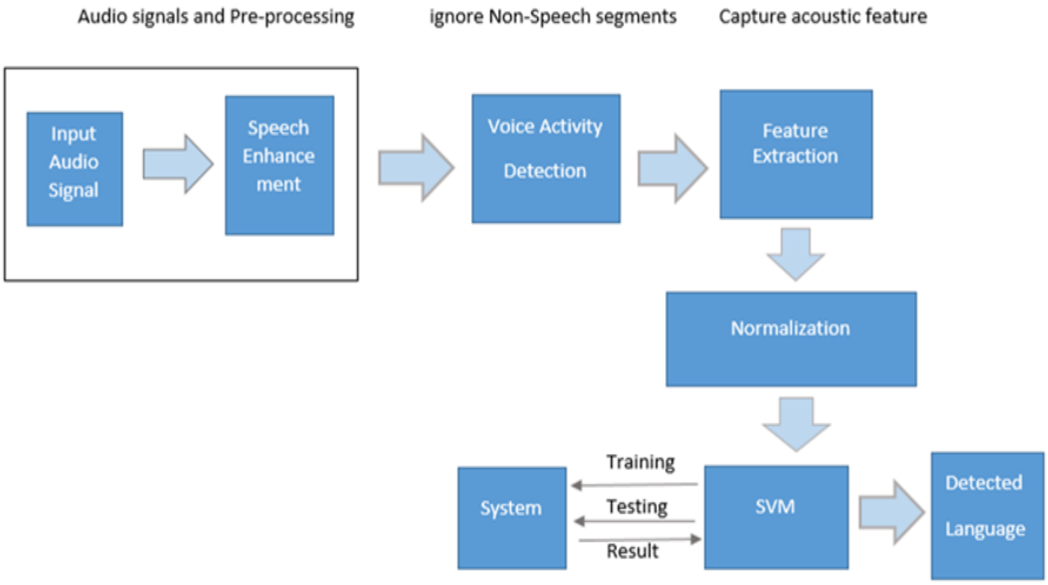

Fig 1: Architecture of Proposed System

5. Train SVM for generated model.

Methodology for Testing phase

Input - Audio signal

Output- Identified language

Processing -

1. Input Signal.

2. Convert to Vocal feature.

3. Apply feature extraction methods. (MFCC, GFCC, PLP, and combination of these feature)

4. Apply UBM model on generated feature

5. Generate I-vector space for UBM model.

6. Supply I-vector to SVM for classification.

7. Classify feature according to trained dataset.

8. Identify Language based on classified feature.

End

4. EXPERIMENTAL RESULTS

4.1 Experimental setup

This system is implemented in .Net Framework 3.5 using c#. To verify the results Indian dataset is used. In this work LID performance is evaluate under the various utterance durations. The database contain two type of dataset, one is Indian dataset with long utterances and one is Indian dataset with short utterances. Long duration utterances dataset contain of 60, 30 or 10 seconds. Short duration utterances dataset contain of 5, 4, 3, 2 seconds. for training phase long utterances dataset are used and for testing phase it consider short utterances.

4.2 Results

The performance of the language identification system can be evaluated in terms of equal error rate (EER). The parameter such as UBM size, dimension of the i-vector are consider.

The Table 1 shows EER on test set of short utterances when the i-vector space is train on matched utterance of long duration. The improve EER is more for short utterances for 1,3,4,5 seconds as compared to long utterances duration of 10,30 seconds. As can be seen from Table 1 EER decreases as the UBM size increases and the performance improvement is more profound when the utterances are shorter.

The Table 2 EER for LID system trained on different Language dataset evaluated on the test sets with short duration. Figure 2,3,4 shows the equal error rate for the system which trained on Hindi, English, and Marathi dataset.

Table 1. Performance in terms of EER (in %) Evaluated on the 1, 3, 4 and 5 seconds test sets.

Test Utterance duration	Duration of Training Utterances					
	1sec	3sec	4sec	5sec	10sec	30sec
1 sec	26.75	-	-	-	20.15	19.15
3 sec	-	15.14	-	-	13.50	11.02
4 sec	-	-	8.75	-	8.35	7.15
5 sec	-	-	-	8.50	8.00	7.10

Table 2. EER of the LID system trained for different languages on short utterances duration test sets.

Language	1sec	3sec	4sec	5sec
Hindi	21.10	11.35	8.75	7.30
English	24.05	10.15	7.38	7.00
Marathi	22.13	13.22	8.08	7.86

Figure 2,3,4 shows the equal error rate for the system which trained on Hindi, English, and Marathi dataset.

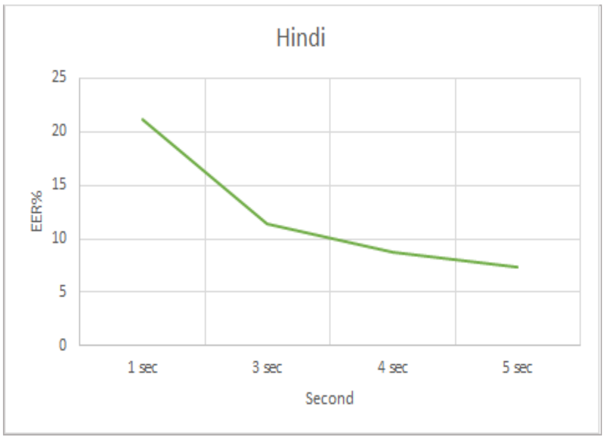

Fig. 2. EER For Hindi Data files

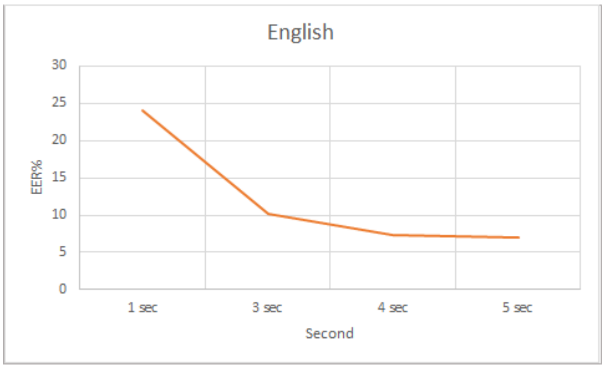

Fig. 3. EER For English Datafiles

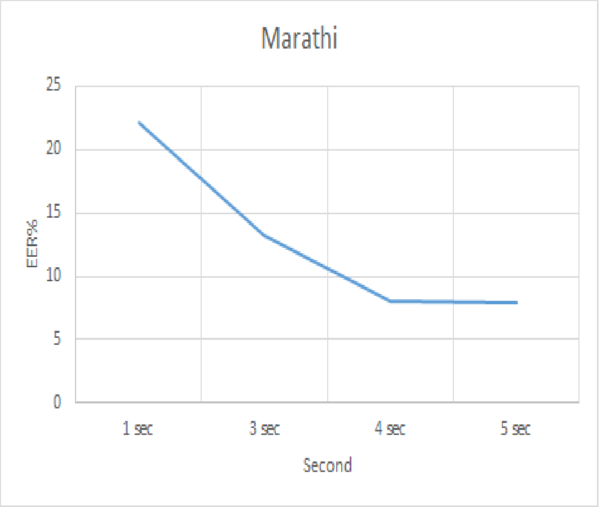

Fig. 4. EER For Marathi Datafiles

5. CONCLUSION

This project present the late mechanical advances in the space of Language Identification. The proposed framework extricates pertinent acoustic elements of the spoken language on the short expressions and deploys into an i-vector based structure. Proposed framework incorporates a novel element representation set, was recommended that brings down the Top error rates when contrasted with standard components. To make quick and practically immediate decision about the spoken language, a simplified i-vector demonstrating structure is used inside of the framework to build the effectiveness of the i-vector extraction process. Methods for precise i-vector space displaying are acquaint with further enhance the distinguishing proof execution on brief length of time discourse articulations.

6. FUTURE SCOPE

Implementation of this technique consider number of features and classification methods for language identification. In future work on language Identification involves system adaptability to changing acoustic environments and integration of phonotactic language information to further lower the error rates on short duration speech utterances.

7. ACKNOWLEDGMENTS

The success and the final outcome of this project required a lot of advice and help from many people and I am extremely fortunate to have got this throughout the end of my project work. I am thankful of my guide, Prof.J.V.Shinde, for her guidance and support. Thanks to all staff members of my college

8. REFERENCES

[1] Y.K. Muthusamy," A Segment Approach to Automatic language identification uses the telephone speech corpus" 1987

[2] Y. K. Muthusamy, N. Jain, and R. A. Cole, "Perceptual benchmarks for automatic language identification," Proc. ICASSP, vol. 1, pp. I–333, 1994..

[3] Y. K. Muthusamy, E. Barnard, and R. A. Cole,"Reviewing automatic language identification," IEEE Signal Process. Mag., vol. 11, no. 4, pp.33–41, Oct. 1994.

[4] M. A. Zissman, "Language identification using phoneme recognition and phonotactic language modeling," in Proc. ICASSP, 1995, vol. 5, pp. 3503–3506.

[5] Y. Yan and E. Barnard, "An approach to automatic language identification based on language-dependent phone recognition," in Proc. ICASSP, 1995, vol. 5, pp. 3511–3514

[6] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," in Proc. ICASSP, 2006, vol. 1, pp. 97–100

[7] E. Singer, P. A. Torres-Carrasquillo, T. P. Gleason, W. M. Campbell, and D. A. Reynolds, "Acoustic, phonetic, and discriminative approaches to automatic language identification," in Proc. Interspeech, 2003

[8] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 788–798, May 2011

[9] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 788–798, May 2011.

[10] A. Larcher, P. Bousquet, K. A. Lee, D. Matrouf, H. Li, and J.-F. Bonastre, "I-vectors in the context of phonetically-constrained short utterances for speaker verification," in Proc. ICASSP, 2012, pp. 4773–4776.]

[11] P. A. Torres-Carrasquillo, D. A. Reynolds, and J. Deller, Jr., “Language identification using gaussian mixture model tokenization,” in Proc. ICASSP, 2002, vol. 1, pp. I–757.

[12] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, “Joint factor analysis versus eigenchannels in speaker recognition,” IEEE Trans.Audio, Speech, Lang. Process., vol 15, no. 4, pp. 1435–1447, May 2007

[13] A. Kanagasundaram, R. Vogt, D. B. Dean, S. Sridharan, and M. W. Mason, “I-vector based speaker recognition on short utterances,” in Proc. Interspeech, 2011, pp. 2341–2344.

[14] A. Larcher, P. Bousquet, K. A. Lee, D. Matrouf, H. Li, and J.-F. Bonastre, “I-vectors in the context of phonetically-constrained short utterances for speaker verification,” in Proc. ICASSP, 2012, pp. 4773–4776.

[15] M.V. Segbroeck, Ruchir Travadi, Shrikanth S. Narayanan,” Rapid language identification,” ieee transactions on audio, speech, and language processing, vol. 23, no. 7, july 2015