IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 38, No. 2, April, 2021, pp. 521-528
Journal homepage: http://iieta.org/journals/ts

HMM Based Language Identification from Speech Utterances of Popular Indic Languages Using Spectral and Prosodic Features

Manchala Sadanandam

CSE, University Engineering College, Kakatiya University, Warangal 506009, Telangana, India

Corresponding Author Email: msadanandam@kakatiya.ac.in

https://doi.org/10.18280/ts.380232

Received: 25 July 2020
Accepted: 10 January 2021

Keywords:
Language Identification System (LID), acoustic features, prosodic features, HMM, Indian spoken languages, pitch and MFCC

ABSTRACT

Language identification system (LID) is a system which automatically recognises the languages of short-term duration of unknown utterance of human beings. It recognises the discriminate features and reveals the language of utterance that belongs to. In this paper, we consider concatenated feature vectors of Mel Frequency Cepstral Coefficients (MFCC) and Pitch for designing LID. We design a reference model one for each language using 14-dimensional feature vectors using Hidden Markov model (HMM) then evaluate against all reference models of listed languages. The likelihood value of test sample feature vectors given in the evaluation is considered to decide the language of unknown utterance of test speech sample. In this paper we consider seven Indian languages for the experimental set up and the performance of system is evaluated. The average performance of the system is 89.31% and 90.63% for three states and four states HMM for 3sec test speech utterances respectively and also it is also observed that the system gives significant results with 3sec test speech for four state HMM even though we follow simple procedure.

1. INTRODUCTION

Language identification (LID) is a sub domain of speech processing technique which identifies the language of spoken utterance by an unknown speaker. It is the task to recognize the language of utterance without knowing the details of speaker and language content. It identifies the languages of speech utterance based on only raw signal of speech utterance [1].

Due to improvements and rapid developments in the technologies and automated applications, language identification plays a prominent role in the automation and globalization demands which deal with multi lingual services and information. LID may be used **in** several applications like call routing, call centres, speech-based automation application, IOT based services in which speech is used, Language translation system etc. [2].

Language identification was also done by considering the log likelihood by considering the average time for some selected speech samples [3]. LID based on Recurrent neural networks by using prosodic features explained [4]. It provides the good accuracy compare to spectral features.

In order to obtain cues, the features of speech are very important. There are four levels features and their cues: Acoustic features (which may be represents by MFCC, SDC). Prosodic features (Pitch, Energy, Duration), Phonetic features (Phoneme recognition), Lexical (words) and Syntax feature [5].

Acoustic features and Prosodic features are obtained from the raw speech signal and these are extracted without knowledge of language, used to design text independent LID like any pattern recognition application [5, 6].

LID also has three major phases having feature extraction, training and testing phase. The methodology of feature vectors extraction and kind of feature vector effects the performance of LID as these feature vectors are input for training and testing phase.

In training phase, generally reference models are created such that one for each language using statistical models like Gaussian Mixture Model (GMM), Discrete Hidden Markov Model (DHMM), Hidden Markov Model (HMM), Neural Networks (NN), Latent Dirichlet Allocation (LDA) etc. These reference models are a compact representation of a huge speech corpus of particular language. In testing phase, the input test speech utterance is labelled with one of known language (represented by reference models) based on the decision criteria. The general procedure to design LID is depicted in the following Figure 1.

The LID approaches are classified into two methods namely: signal-based system and text-based system. Spectral features and prosody features are come under signal bases system to design LID whereas phone recognition, word level recognition and continuous approaches are come under text-based system. In speech signal-based approach, the feature vectors are extracted from raw signal without knowing details of language and content of speech. In our paper we follow this approach to design LID to identify Indian languages using acoustic features i.e. MFCC and Prosody feature i.e. Pitch.

Any efficient LID must able to get information of speech utterances which discriminate different languages from a large speech corpus and also it must be flexible to increase the number of languages. It is very essential to extract some cues for designing LID in order to get knowledge on speech information and relationships among cues. Human beings can do easily recognize the language of speech utterance after hearing, if they know the knowledge of the languages in terms of phoneme or words. For machines also, speech and its phonemes also very important and essential for designing an

LID. The phoneme is smallest unit of speech in a spoken language [7].

With the literature survey, I have learnt that there are several statistical and classify models like Vector Quantization, GMM, HMM, SVM, LDA, fuzzy set approaches and rough set approaches etc. Hidden Markov Model is popular statistical model to implement speech processing tasks.

1.1 Applications and challenges of LID

LID is trained only once for one language and requires shorter time to recognise test utterances and LID can run simultaneously on multiple system so that LID can be used in several applications like in call centres to route the call to language specific server based on the language of caller voice, in defends global terrorism, to suspect the terrorist. These are also used in Human Computer Interaction (HCI), Speaker Identification and biometric voice applications.

The major challenging issues to deal with LID are speaker characteristics variations, variation in accents, differences in dialects and there exists similarities in languages (in case of Indian languages have common root words and similar grammatical structure so it is challenging task) as there are no proper method or technique available for speech processing. In Indian languages, as there all root words & most of languages have same origin even though few unique words, it is very difficult to implement.

In this paper, we proposed an LID system which is designed using HMM and hybrid features with the combinations of acoustic features like MFCC and Prosodic features like Pitch. We carried out experiments on Indian language database and accuracy of the system is significant.

The paper is organised into five sections. Section 1 deals with introduction and literature survey of the paper including application and challenges of LID. Section 2 describes Literature survey, Section 3 describes the feature extracted methods, Section 4 elaborates about Hidden Markov Model and its working principle. Section 5 gives about proposed methodology including training and testing with HMM and section 6 is about experiments and results. Finally, conclusion is given in section 7.

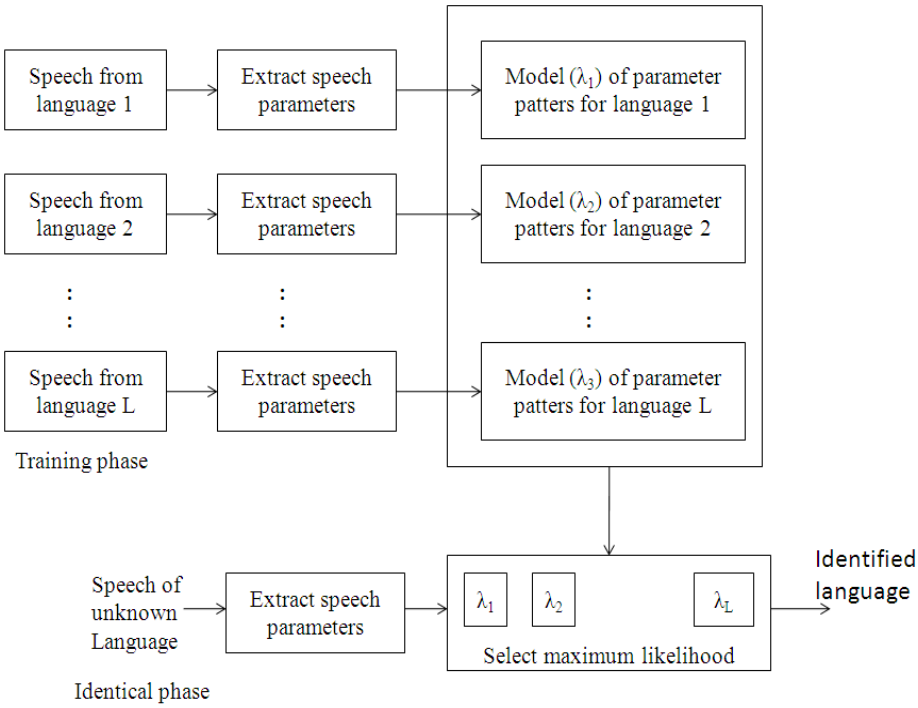

Figure 1. General framework of Language Identification System (LID)

2. LITERATURE SURVEY

The first LID was designed and explored with acoustic features to build a dictionary of words using speech sound. Prosody features also play a vital role in recognizing language from speech signal [1, 2]. Prosodic features like pitch, energy, stress are different in tonal language compared to non-tonal languages [3, 4]. Adeeba and Hussain [3] used BLSTM method to identify the Urdu language with short duration of speech signal. The duration of speech signal is an average 8s. They applied BLSTM method with different feature extraction techniques like MFCC and GFCC. The performance of the system is compared to that MFCC is providing good accuracy rather than GFCC. In this they extracted acoustic features from speech signals to identify the language.

This paper [7] has shown that bilingual code-switched speech: the speaker speaks primarily the host language but with a pre identified different combinations language like English-Hindi, English-Bengali and vice versa by using two models of Recursive Neural Network techniques i.e. Long Short-Term Memory (LSTM) and bidirectional LSTM and compare with Conditional Random Fields (CRF) classifier. They observed that RNN models produced good accuracy compare to CRF. Heracleous et al. [8] applied various deep learning models like DNN, CNN to LID. Authors compare the results with baseline method SVM and observed there is a significant improvement in performance of deep learning models.

Jamatia et al. [9] identified different combinations language like English-Hindi, English-Bengali and vice versa by using

two models of Recursive Neural Network techniques i.e. Long Short-Term Memory (LSTM) and bidirectional LSTM and compare with Conditional Random Fields (CRF) classifier. They observed that RNN models produced good accuracy compare to CRF.

Tang et al. [10] identified languages using acoustic level feature. They combined three different methods like HMM states and Gaussians to for this purpose. They got good accuracy such imbalanced data with these models compare to general methods.

The authors [11, 12] performed LID task using vector quantization with spectral features on the OGI database. Kirchhoff et al. [13] applied the n-gram method to LID by using phonatactic features. Nagarajan designed the LID task on the OGI database using the frequency acoustic feature vectors in languages and achieved 59.5% accuracy with VQ and GMM [14, 15].

Salamea et al. [16] used new approaches for language identification. They used RNN and Neural Embeddings for designing LID. Neural embedding is used for the feature extraction after that applied the RNN to identify the language. In this they used phono grams instead of phonetic features and they increased the accuracy 23.0% in identification on KALAKA-3 database. They also applied MFCC acoustic in vector and PPRLAM model and they got 39.3% improvement.

Zhang [17] used an unsupervised deep learning model to identify the language and dialect identification. He used 4-way dialects of Chinese language for this purpose. They extracted unsupervised features and trained with phonetic labels after that auto encoder model to process the features extracted. The authors implemented LID using pitch and syllable time and achieved good results in the study [18] and compare the different approaches on telephonic speech to LID [19].

The authors designed LID using spectral features and their frequency with VQ, GMM and HMM for OGI database languages and some Indian languages [20, 21].

Ma et al. [22] used small speech utterances for intelligent vehicle controlling. The controlling of vehicle almost requires small speech sample but small speech signals not enough to extract the required features. So, authors used LSTM to extract the enough features from speech. They trained the speech feature but using LSTM NN model. To increase the accuracy and remove the unvoiced speech they used TSM method on AP17-OLR database. Bartz et al. [23] designed using convolution recurrent neural networks (CRNN) and the performance is compared with different approaches.

3. FEATURE EXTRACTION

Feature extraction is a major step to recognition of any pattern like language, dialect, speaker from speech utterances etc. from speech signal. The performance of proposed system depends on feature vectors, the selection of feature vectors and some other parameters of features are very important to get good and significant results.

In case of language identification, the selection of feature vectors must discriminate the content of speech i.e. phoneme, sequence of phoneme and frequency of phoneme and as well as energy and pitch of vocal tract.

According research survey, the frequency of phoneme is different in Indian language is different and also pitch and energy of speech signal of Indian language speaker may vary.

They are several LID cues which discriminate the phonemes among Indian languages. Among those cues, Mel Frequency Cepstral Coefficients are important feature of speech signal which reveals phonetic differences among languages. Pitch and energy also play vital role to distinguish the language. In our work, we extract the MFCC and Pitch from short term duration of windowed speech signal.

3.1 Mel frequency cepstral coefficients

MFCC are popular acoustic features and these have significant results in speech processing tasks. These features mainly extracted from pre processed speech signal. The steps to extract MFCC from speech signal are described in Figure 2.

The extraction of MFCC feature vectors from speech signal follows six step procedures [20]:

First step is pre processing in which signal is smoothen by removing noise using digital filter (pre-emphasis filter) in order to enhance the performance of efficiency of system.

The second step is framing and windowing, instead of analyzing the entire speech signal at once signal, it is split up into overlapped frames with short time duration, generally frame size is 10ms -30ms. The information at beginning and ending of frame is very important. To avoid the loss of information, we overlap the frames to preserve the information. Windowing technique is applied to avoid discontinues in signal so that Hamming window is applied on speech signal.

In third step, apply fast Fourier transformation to obtain the magnitude frequency of each frame of windowed speech signal.

In fourth step, Using Mel scaled filter bank, smoothen spectrum of speech signal which obtain the data values of spectrum from more significant parts Mel frequency scale is linear frequency which is below 1000Hz space and logarithmic space above 1000Hz.

Mel scale is defined as Eq. (1),

$$Mel(f)=2595*\log10(1+f/700) \quad (1)$$

where, f is a frequency in Hz.

In fifth step, Logarithm is applied to Mel spectrum which converts Mel spectrum to time domain i.e. Mel cepstrum, then apply Discrete Cosine Transform (DCT) to Mel cepstrum to obtain the coefficient.

The sixth step reduces co-relation between compressed information and coefficients into lower order coefficients. MFCCs is calculated using the Eq. (2)

$$C_n = \sum_{k=1}^{K}(\log S_k)[n(k-(1)/(2))(\pi)/(K)] \quad (2)$$

where, k is typically chosen as 10-15, S_k Spectrum of speech Signal, C_0 is excluded since it represents mean value of input signal which denotes speaker specific information.

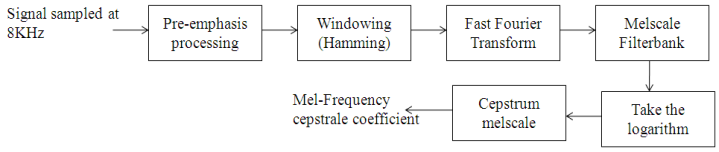

Figure 2. Extraction of MFCC features vectors

3.2 Prosodic features

Prosodic features are also important feature to discriminate macro level information such as phoneme information and relations among languages. Prosodic feature are pitch contour rhythm, stress and fundamental frequency (f0) play vital role in speech recognition. Generally prosodic features are used combination with acoustic features to improve accuracy of LID systems as these MFCC feature vectors carry the information about phonemes and discriminate occurrence of frequency of phonemes among languages. MFCC feature vectors are best features involved to design language identification system but some Indian languages like Telugu, Pitch and energy are also show more significant variation with other languages so that pitch and energy are suitable features to classify the Indian languages in order to increase the accuracy of system [10]. We choose 13 dimensional MFCC and 1-pitch as feature vectors and concatenated to form hybrid feature vectors with 14 dimensionalities.

3.2.1 Pitch

Pitch is also called as fundamental frequency of speech signal. It is used to identify the movement of glottis and its characteristics. The highest or low frequency of speech of signal when the vibration in glottis is happened called as Pitch. It represents physical characteristics of vocal cord used to easily identify the speaker or voice.

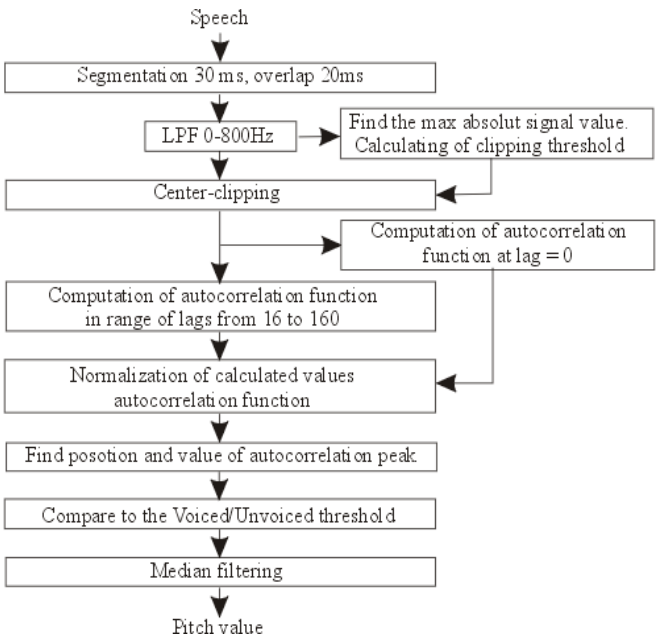

Figure 3. Different steps involved in pitch calculation

In this work, we used Simplified Inverse Filtering Technique (SIFT) method to extract the pitch from speech utterances. It is depending upon Linear Prediction (LP) estimation from speech signal. The SIFT method works like autocorrelation method internally for extracting pitch but difference between them is SIFT applied autocorrelation on LP residual instead of applied on speech signal directly. LP residual calculated from LP coefficients where the LP coefficients carry the vocal track information. So, LP residual contains essential information and also there is no ambiguous peaks represents the same pitch in the time period T0. In case of LP analysis, the vocal tract information is modelled in terms of LP coefficient (LPC) as the process of the prediction of current sample which is a combination of 'p' samples. The LPCs usually represent the coefficients of the LP filter which in turn is the representation of vocal tract. And by using the "Inverse Property" corresponding inverse filters are constructed through which when speech signals are passed, results in LP residual as an output. SIFT method clearly discriminate the voiced and unvoiced speeches in order to find the pitch values. Basic structure of pitch extraction using SIFT as shown in below Figure 3.

Firstly, it takes the speech as input and applies the pre-processing on the speech in order to remove the noise and unvoiced portion from speech utterance. In this method we applied digital filter for pre-processing purpose. After the pre-processing we divide the speech signal by apply the hamming window techniques like each segment of speech 30ms with overlap of 20ms in order to reduce the changes in features. After applying the filtering applied the SIFT method to extract the peak levels in the speech utterances i.e. pitch value of speech signal. The basic model as shown in the Figure 4.

Figure 4. Basic model to extract the pitch

3.2.2 Hybrid features

In this paper, we concatenated 13 dimensional MFCC feature vectors and 1 dimensional prosodic feature i.e. pitch to form the 14-dimensional feature vectors for the Language Identification. As the MFCC feature vectors independently do not discriminate the cues for some Indian language like Telugu, we Consider prosodic feature which discriminate the pitch. The prosodic feature provides the exact information about the vocal tract shape and length. In this, we used the prosodic feature i.e., Pitch to clearly discriminate the words from different languages. In 13 dimensional MFCC features, 1st feature indicates the fundamental frequency of the speech utterance and remaining 12 features are Cepstrum coefficients of short-term windowed speech signal. These features include the details regarding the rate variations in the different frequency bands.

4. HIDDEN MARKOV MODEL

Hidden Markov Model (HMM) is double stochastic models which are more familiar and popular in speech processing tasks. As these well analyze the spectral properties of speech signal, observe and catch the durational and temporal variants among phonemes /speech sounds, are very suitable for speech processing tasks including language identification [20]. These models take the sequence of feature vectors as input and return likelihood value of feature vectors for unknown speech utterance.

HMM comprises mainly three components namely states, transition probabilities in states and probability of state symbol. HMM is represented mathematically by three parameter i.e. $\lambda=(\prod,A,B)$ where $\prod$ is set of initial probabilities of states i.e. $\prod=\{\prod_i\}$, A is matrix which denotes state transition between two states and B is observation symbol probability distribution in each state of B. If the observation sequence is discreet value, it is called discrete Hidden Markov model otherwise if observation sequence is continuous, it is called Continuous

Hidden Markov model.

In Continuous HMM, B consists of Mixture of Gaussian functions and $B=b_j(O_t)$ where:

$$b_j(O_t) = \sum_{m=1}^{m} C_{jm} \aleph\left(\mu_{jm}, \sum_{jm}, O_t\right) = \frac{1}{(2\pi)^{D/2}\left|\Sigma_{jm}\right|^{1/2}} exp\left[-\frac{1}{2}(x-\mu_{jm})^T \sum_{jm}^{-1}(x-\mu_{jm})\right] \quad (3)$$

where, c_{jm}=weighting coefficient, μ_{jm}=mean vectors, $\sum jm$=covariance matrices c_{jm} should satisfy the stochastic constraints $c_{jm}\geq 0$, $1\leq j\leq N$, $1\leq m\leq M$.

Based on the transition property between states of HMM, HMMs are characterized two types namely left-right HMM and Continues Ergodic HMM. In left-right HMM, state transition takes place from left state to right state such that no transition is permitted to sate with lower index to current state index. The following diagram describes the left-right state transition in Figure 5.

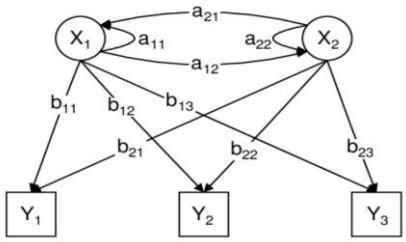

Figure 5. Transition diagram for left-right Hidden Markov Model

Ergodic HMM is fully connected and it allows transitions between any two states. So that ergodic HMM captures positional and temporal pattern effectively with all possible combinations of all sounds. With advantages of ergodic HMM structure, it more popular and gives significant results in speech processing techniques including LID. Figure 6 describes the state transition procedure.

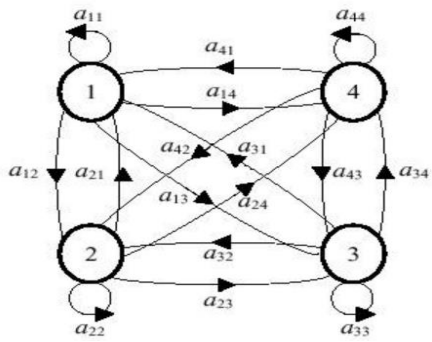

Figure 6. Transition diagram for Ergodic Hidden Markov Model

HMMs are designed and analyzed with three its associated problems. These Associated problems are:

1. Evaluation problem
2. Decision problem
3. Optimization problem

Evaluation problem deals with evaluation of probability/likelihood value of observation sequence against given an HMM. With this problem, testing is performed with forward-backward algorithm [12].

Decision problem computes appropriate sequence of states for given HMM and the probabilities. Optimization problem is used to design HMM and optimize the parameters ∏, A, B of HMM using EM algorithm [20]. This problem is used to train HMM to give compact representation of huge speech corpus.

With advantages of HMM, we use HMM to create reference model for each listed language in our paper and designed HMMs with different number of states like three, four.

5. PROPOSED METHOD

The proposed method implements text independent language recognition system using Hybrid feature vectors which are extracted from speech corpus of different Indian language This methodology has two step procedure such that training phase and testing phase.

In training phase, the reference models are created for each language using Hidden Markov models where mean vectors, covariance matrix and mixture weights are parameters of hidden Markov model. In testing phase, language of short duration of unknown speech is recognized by evaluating the feature vectors of test samples with n number of HMMs where n is no of languages considered for implementation of system.

5.1 Training phase

Training phase of this methodology is two step procedure in first step, the feature vectors of training speech of all consider Indian languages are extracted by applying windowing technique using the method described in section 2. Each language consists 25 minutes duration of speech as training speech corpus. We extract 12 dimensional MFCC feature vectors and Pitch and Concatenated so that 14-dimensional feature vectors from each listed language. The language specific 14-dimensional feature vectors are used to training a language specific HMM. Hidden Markov Models are designed as one reference model for each considered language using re-estimation algorithm [20] as described in section 3. Training procedure diagrammatically in the following Figure 7.

5.2 Testing phase

We follow three step procedures for testing phase of proposed method. The abstract flow diagram of testing phase is depicted in Figure 8. First step of testing phase is feature extraction in which 12 dimensional MFCC and pitch are extracted and concatenated to form 14-dimensional feature vectors from short time windowing unknown speech of test samples as described in section 2. The sequence of 14-dimensional feature vectors is chosen as the observation sequence of Hidden Markov Model. In second step, the feature vectors of unknown test speech utterance are applied for the evaluation against each language specific HMM in order to compute likelihood value of each feature vector of unknown utterance. Third step calculates the mean value of likelihood of each feature vectors of unknown test samples and reveals the language of unknown utterance of speech which HMM gives max mean value of likelihood.

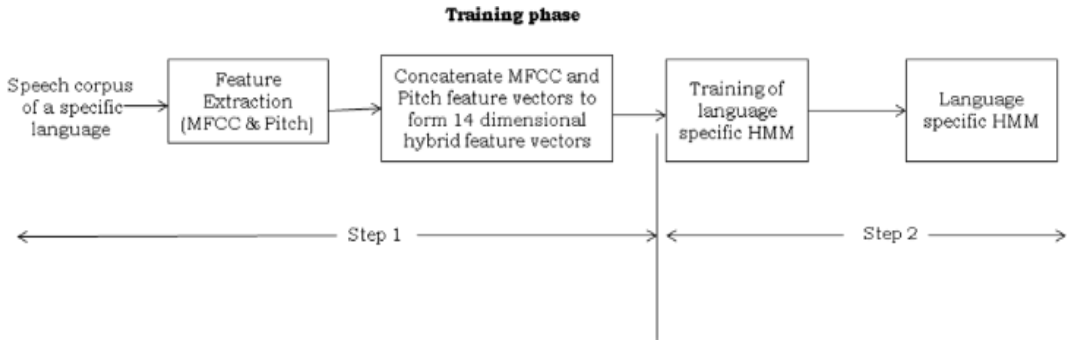

Figure 7. Training phase of LID using HMM

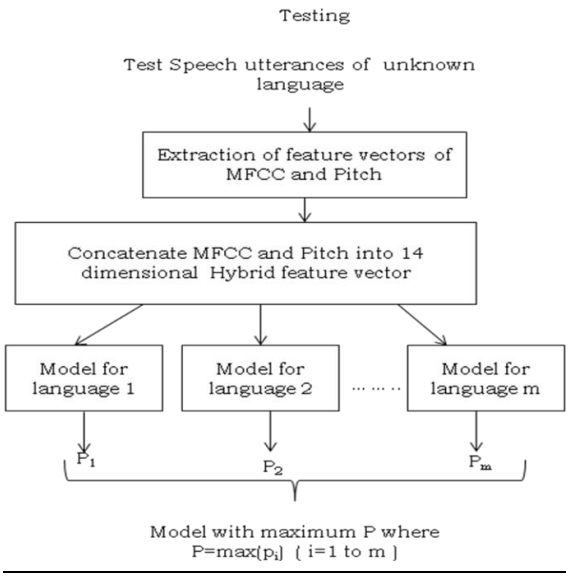

Figure 8. Testing phase of LID using HMM

6. EXPERIMENTS SET UP AND RESULTS

We carried experiments on Indian languages data base [24, 25] using Python on Windows platform. We consider nine popular Indian languages and took 20 mints of speech of each language for training to create reference model for each language. We tested 100 test sample with different durations like 1sec, 2sec and 3sec. Continues Hidden Markov model with three states and four states are trained using 14-dimensional feature vectors of each language. Testing is carried out for different duration's speech utterances of considered languages of 1sec, 2sec and 3sec using proposed method. The performance of proposed system for Indian languages for various duration of test speech utterance is depicted in the following tables with MFCC and pitch concatenated feature vectors with different states of HMM. Table 1 is the results of proposed method for different Indic Languages for different test durations of speech with three states of HMM and Table 2 represents the results of system with four states of HMM.

From the results tables, it is observed that the average performance for the system is more with four state HMM compare to three state HMM and also observer system performs good at 3sec test speech duration and the corresponding graphs for different no of states of HMM and different test duration of speech is given below for Indic Languages in Figure 9 and Figure 10.

Table 1. The results of proposed model with three state HMM and feature vectors of MFCC and Pitch

Language	Performance in % for 3 state HMM		
	1sec	2sec	3sec
Tamil	96.2	97.5	97.3
Oriya	76.5	77.3	79.1
Hindi	89.4	91.4	94.8
Guajarati	80.1	81.2	84.9
Telugu	86.2	86.3	88.4
Marathi	84.2	87.7	88.4
Assamese	77.3	77.7	81.3
Malayalam	93.2	95.5	95.4
Kannada	91.5	93.6	94.2
Average	86.07	87.58	89.31

Table 2. The results of proposed model with four state HMM and feature vectors of MFCC and pitch

Language	Performance in % for 4 state HMM		
	1sec	2sec	3sec
Tamil	96.5	98.2	99.1
Oriya	78.6	79.2	79.3
Hindi	89.3	90.2	95.4
Guajarati	81.2	85.3	87.2
Telugu	87.7	89.2	89.4
Marathi	85.4	87.8	90.2
Assamese	78.1	78.3	81.6
Malayalam	92.4	94.6	96.1
Kannada	92.9	94.2	97.4
Average	86.9	88.56	90.63

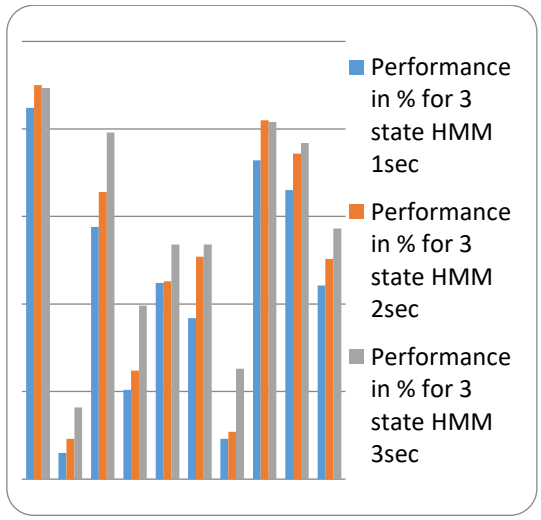

Figure 9. Performance of LID with 3 states HMM for 1sec, 2sec and 3sec

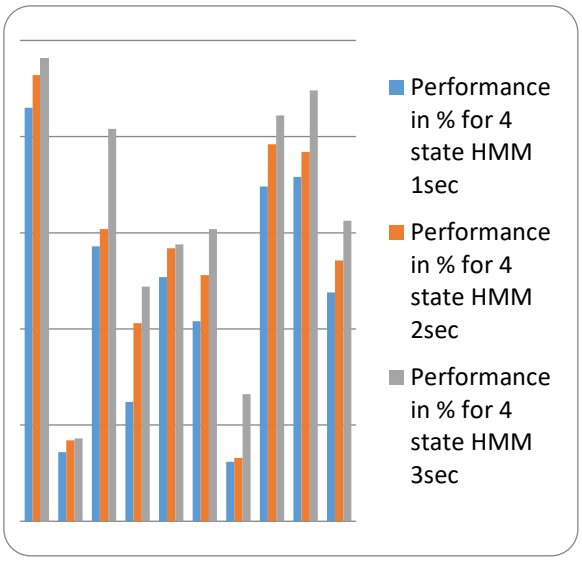

Figure 10. Performance of LID with 4 states HMM for 1sec, 2sec and 3sec

7. CONCLUSIONS

In this paper we have designed HMM based text independent language recognition system to identify the language of unknown utterance of speech in Indian languages. We have extracted 14-dimensional feature vectors which are formed by concatenating MFCC and pitch from the speech signals. We have carried our experiments to design HMM with various states and tested system for different durations of test speech. The average performance of the system is 89.31% and 90.63% for three states and four states HMM for 3sec test speech utterances respectively and also it is also observed that the system significant results with 3sec test speech for four state HMM.

REFERENCES

[1] Kotsakis, R., Matsiola, M., Kalliris, G., Dimoulas, C. (2020). Investigation of spoken-language detection and classification in broadcasted audio content. Information, 11(4): 211. https://doi.org/10.3390/info11040211

[2] Leonard, R.G., Doddington, G.R. (1974). Automatic Language Identification. Technical Report RADC-TR-74-200, Air Force Rome Air Development Center.

[3] Adeeba, F., Hussain, S. (2019). Native language identification in very short utterances using bidirectional long short-term memory network. IEEE Access, 7: 17098-17110. https://doi.org/10.1109/ACCESS.2019.2896453

[4] Cummins, F., Gers, F., Schmidhuber, J. (1999). Language identification from prosody without explicit features. Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99), pp. 371-374.

[5] Savic, M., Acosta, E., Gupta, S.K. (1991). An automatic language identification system. [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, Toronto, ON, Canada, pp. 817-820. https://doi.org/10.1109/ICASSP.1991.150462

[6] Sugiyama, M. (1991). Automatic language recognition using acoustic features. [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing, Toronto, ON, Canada, pp. 813-816. https://doi.org/10.1109/ICASSP.1991.150461

[7] Yeh, C., Lee, L. (2015). An improved framework for recognizing highly imbalanced bilingual code-switched lectures with cross-language acoustic modeling and frame-level language identification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(7): 1144-1159. https://doi.org/10.1109/TASLP.2015.2425214

[8] Heracleous, P., Takai, K., Yasuda, K., Mohammad, Y., Yoneyama, A. (2018). Comparative study on spoken language identification based on deep learning. 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, pp. 2265-2269. https://doi.org/10.23919/EUSIPCO.2018.8553347

[9] Jamatia, A., Das, A., Gambäck, B. (2019). Deep learning-based language identification in English-Hindi-Bengali code-mixed social media corpora. Journal of Intelligent Systems, 28(3): 399-408. https://doi.org/10.1515/jisys-2017-0440

[10] Tang, Z., Wang, D., Chen, Y., Li, L., Abel, A. (2018). Phonetic temporal neural model for language identification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(1): 134-144. https://doi.org/10.1109/TASLP.2017.2764271

[11] Gianni, L., Frederking, R., Minker, W. The basic architecture of a language id. CMU, 1996. https://www.cs.cmu.edu/~ref/mlim/chapter7.html, accessed on 20 October 2020.

[12] Thyme-Gobbel, A.E., Hutchins, S.E. (1996). On using prosodic cues in automatic language identification. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, Philadelphia, PA, USA, pp. 1768-1771. https://doi.org/10.1109/ICSLP.1996.607971

[13] Kirchhoff, K., Parandekar, S., Bilmes, J. (2002). Mixed-memory Markov models for Automatic Language Identification. 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, pp. I-761-I-764. https://doi.org/10.1109/ICASSP.2002.5743829

[14] Nagarajan, T., Murthy, H.A. (2002). Language identification using spectral vector distribution across languages. In Proc. International Conference on Natural Language Processing, pp. 327-335.

[15] Nagarajan, T., Murthy, H.A. (2004). Language identification using parallel syllable-like unit recognition. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. I-401. https://doi.org/10.1109/ICASSP.2004.1326007

[16] Salamea, P., Christian, R., D'Haro, L.F., Cordoba, R. (2018). Language recognition using neural phone embeddings and RNNLMs. IEEE Latin America Transactions, 16(7): 2033-2039.

[17] Zhang, Q., Hansen, J.H.L. (2018). Language/dialect recognition based on unsupervised deep learning. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(5): 873-882. https://doi.org/10.1109/TASLP.2018.2797420

[18] Zissman, M.A. (1993). Automatic language identification using Gaussian mixture and hidden Markov models. 1993 IEEE International Conference on

Acoustics, Speech, and Signal Processing, Minneapolis, MN, USA, pp. 399-402. https://doi.org/10.1109/ICASSP.1993.319323

[19] Zissman, M.A. (1999). Comparison of four approaches to automatic language identification of telephone speech. IEEE Transactions on Speech and Audio Processing, 4(1): 31-44. https://doi.org/10.1109/TSA.1996.481450

[20] Rabinar, L.R. (2013). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2): 257-286. https://doi.org/10.1109/5.18626

[21] Singer, E., Torres-Carrasquillo, P.A., Gleason, T.P., Campbell, W.M., Reynolds, D.A. (2003). Acoustic, phonetic, and discriminative approaches to automatic language identification. Proc. EUROSPEECH 2003, pp. 1345-1348.

[22] Ma, Z., Yu, H., Chen, W., Guo, J. (2019). Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features. IEEE Transactions on Vehicular Technology, 68(1): 121-128. https://doi.org/10.1109/TVT.2018.2879361

[23] Bartz, C., Herold, T., Yang, H., Meinel, C. (2017) Language identification using deep convolutional recurrent neural networks. In: Liu D., Xie S., Li Y., Zhao D., El-Alfy ES. (eds) Neural Information Processing. ICONIP 2017. Lecture Notes in Computer Science, vol 10639. Springer, Cham. https://doi.org/10.1007/978-3-319-70136-3_93

[24] Language Technology Research Center, IIIT Hyderabad. https://irel.iiit.ac.in/uploads/README_0, accessed on 20 October 2020.

[25] Haris, B.C., Pradhan, G., Misra, A., Prasanna, S.R.M., Das, R.K., Sinha, R. (2011). Multi-variability speech database, speech lab. Indian Institute of Technology Guwahati, Guwahati. https://doi.org/10.1007/s10772-012-9140-x