International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 12, December-2014

A Novel Language Identification System For Identifying Hindi, Chhattisgarhi and English Spoken Language

Khagesh Kumar Sahu[1]
M. E. Scholar, SSCET Bhilai

Mr. Vinay Jain[2]
Associate Professor, E&TC, SSCET Bhilai

Abstract-: **Different languages are spoken by different people in different part of the world. It is very necessary to design some system which can identify the spoken language i.e. Language identification system (LID). This paper present an efficient language identification system using MFCC feature of speech and feed forward neural network as a classifier. This system is designed to identify languages Hindi, English and Chhattisgarhi. This system can also be used for any other languages.**

Keywords- MFCC, LPC, Feed forward Network ,HMM

I INTRODUCTION

Though the globalization has made all the people close to each other but there is still hindrance in global communication due to the lack of common communication medium. Different people in different part of globe speak different language. In order to effective communication there is a requirement of such language which is known to both the parties. Language identification(LID) system offers the solution to this problem. Once the language is identified automatically then it can be converted into other language using some add on software. Language identification is basically a speech processing operation which can be used for translating spoken language to any other known language. It is also used in multilingual speech recognition [1]. It also find application in spoken document retrieval [2].

II BACKGROUND

In the past a lot of research work has been done in this field and in the last decades this field witness a significant progress. For a human being, it is easy to identify the spoken language if the human being is familiar with the spoken language. But for the machines or computer, it is very difficult to identify the spoken language. Language identification system takes the speech of any language as input and with the help of some features of the speech signal, identify the language. Morphology, Prosody, phonology and syntax are some of the characteristics that makes one language differ from other.

The first significant effort in language identification was done by Texas Instrument [3]. This work was based on the fact that the frequency occurrences of some specific sound are different in different language and if these are extracted out then languages can be identified. Later on with the help of human interactive approach was also incorporate [4] to increase the identification accuracy. But this method fails to give good accuracy as we increase the number of languages. In 1977 a manual phonetic transcribed database method was proposed [5]. This approach first extract out phonetic labels from phonetic transcription then trained HMM using these labels. In 1980[6], it was proposed that if HMM model is applied to real speech data, better performance can be achieved. In 1982[7], acoustic feature like log area ratio, filter coefficient , cepstral coefficient and formants frequencies based method was proposed in language identification method. This method achieved 84% accuracy.

In 1986[8] an LID system was proposed which was based on extracting the formant frequency from the speech and then classifying it with the help of k-means or vector quantization method.

In 1989, goodman[9], proposed a modification on the work proposed by foil[8].He showed that by adding more features of the speech and improvement in the classifier can give much better results.

Sugiyama in 1991[10] proposed a LPC (Linear prediction coefficient) based language identification system. In his system, vector quantization was used for classification.

Nakagawa in 1992[11], presented four different method i.e. Vector quantization(VQ), Continuous density HMM , Discrete HMM and GMM(Gaussian Mixture distribution model).

Muthusamy[12] in his dissertation suggested that broad phonetics , prosodic information along with acoustic information is required for automatic language identification.

In 1995 Yan[13] in his dissertation studied the role of acoustic, phonotactic and prosodic information for language identification system. He also introduced two model in his dissertation i.e. backward LM and context dependent duration model. He achieved 91% accuracy in 45 second segments and 77% accuracy in 10 second segment.

In 1996 schultz[14] presented a language identification system which was based on vocabulary speech recognition system. In his model he adopted phone lavel and word level based identification with or without language model

Berkling in 1999[15] presented various approaches for computing the confidence measure for LID system. He proposed three different types of confidences- first type of confidence scores all poles according to the winner.

In this paper, language identification method which was based on MFCC(Mel Frequency Cepstral Coefficient) feature of the spoken speech is presented. Feed forward neural network is used for classification purpose in this method.

III PROPOSED METHOD

In any language identification algorithm, selection and extraction of appropriate features from the speech signal play very important role. Selected feature must be able to reflect, detect and differentiate the variation in speech of different language. In this algorithm, three features i.e. MFCC, Delta coefficient and double delta coefficient are used as feature.

Block diagram of proposed language identification system is shown in figure given below. First of all the speech signal of different languages are acquired through the mike and computer system and stored in a memory. Then each signal is read using MATLAB command and then a low pass filter is applied to each speech signal to get rid of high frequency signal. Another reason for applying low pass filtering is to make the speech signal band-limited so that aliasing effect can be avoided. Next step is to divide the speech signal into different frames of 25ms length. Speech signal itself is very lengthy signal and it is very difficult to extract out meaningful features from the full speech signal therefore speech signal is divided into frames of some predefined size. The size of frame must be lengthy enough to contain all the necessary information or feature. Frame size of length 25ms or 20ms are generally used. Once the speech signal is divided into different frames then feature extraction block extract the feature from each frames. In the proposed method MFCC coefficient, Delta coefficient and double delta coefficients are used as features. All the three coefficients from each frames are extracted and stored in a variable which act as a feature vector for this method. In this way database of speech features are created and stored.

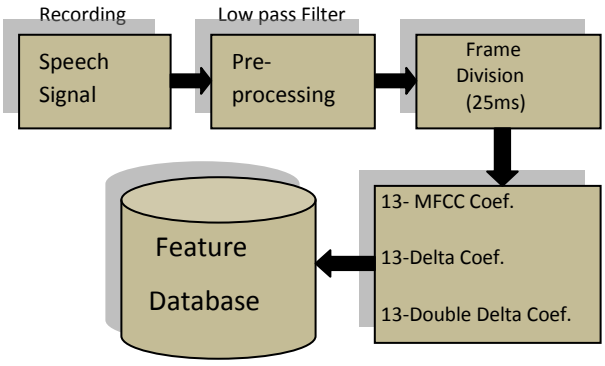

Figure1 Block Diagram

Since the clips of most of the speech samples are from four to seven second long which can increase the computational time therefore the middle part of 1.5 second long is cut out from each clip and then all the features are extracted from theses 1.5 second long speech samples. Each 1.5 second clip is then divided into 25 ms frames with overlapping distance of 10ms. Each frame is then multiply with hamming window for smoothening the discontinuities present in the speech signal. Mel-frequency cepstral coefficients are then extracted from these frames. Mel-frequency cepstral coefficient represent the audio signal in mel scale or in other words it maps the audio frequency nonlinearly just like a human ear. Generally 13 mfcc coefficients of each speech frames are sufficient to represent the speech property of frames. Beside MFCC, we also incorporated the delta features (i.e delta coefficients and double delta coefficients).

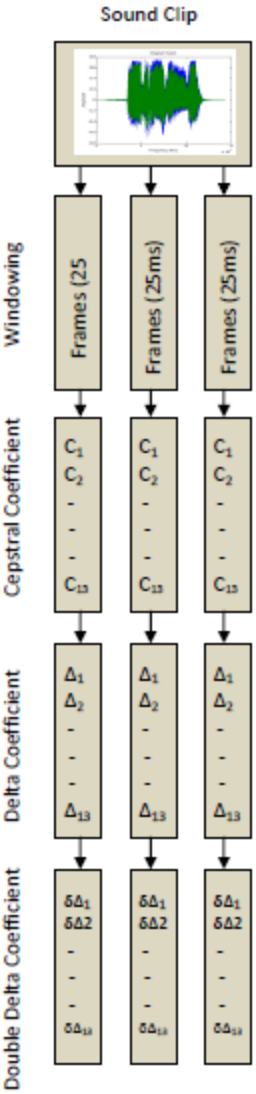

Figure 2 Feature Extraction

Delta and double delta coefficients are simply first and second order derivatives of cepstral coefficients. Delta features captures the variation in cepstral coefficients which are essential for language classification. Therefore 13- delta and 13- double delta coefficients are also

extracted out from each frames. This makes total 39 features from each 25ms frame. Since in our approach we are going to use neural network for classification purpose, we adopted a different approach for feature extraction. Therefore in our approach, we extracted out the features from each clip rather than from each frames of 25ms. From each frames of each clip, we extracted out 13 MFCC, 13-Delta coefficients and 13-Double Delta coefficients and then we took mean and variance of each of these feature across each clip making 78-different features (i.e. 13-mean MFCC,13-meanDC,13-meanDDC and 13-varianceMFCC,13-varianceDC, 13-varianceDDC) from each clips. This approach is good as it averaging out noise from the speech signal and also reduces the dimension of feature vectors. In this work, 200 sample for each language has been taken (i.e. 200 for Hindi, 200 for English and 200 for Chhattisgarhi).Therefore the dimension of feature vector for each language is of 200x78 size making total size of feature vectror for all the language is of size 600x78. A target vector is also created of dimension 3x600 (since there are three languages). 70% of feature vector is used as training data while 30% of feature vector is used for testing data.

Once the database of speech feature is created then the next phase is to design appropriate neural network which can be trained by theses extracted feature. In this algorithm, Feed forward neural network has been chosen for classification purpose. A feed forward neural network having 3-layers of neurons has been designed. first layer is a input layer (IL), second layer is a hidden layer(HL) and the output layer(OL). Input layer (IL) has 78 neurons (i.e. one for each feature) while the hidden layer comprises of 10-neurons and the output layer has 3-variable (one for each language). Neural network design is shown in figure given below. Each neuron of input layer is connected to each neuron of hidden layer while the each neuron of hidden layer goes to the each neuron of output layer. In the given neural network model "s" is sigmoid function given by-

$$s(x) = \frac{1}{1+e^{-x}}$$

b_i^{HL} = bias of neuron i in hidden layer(HL)

$w_{j,i}^{HL}$ = weight of every neuron i in Hidden layer(HL) for each input j.

Once the neural network is designed then with the help of training feature vector data and target vector, neural network is trained with the help of backward propagation method. Once the neural network is trained then it is tested using testing data.

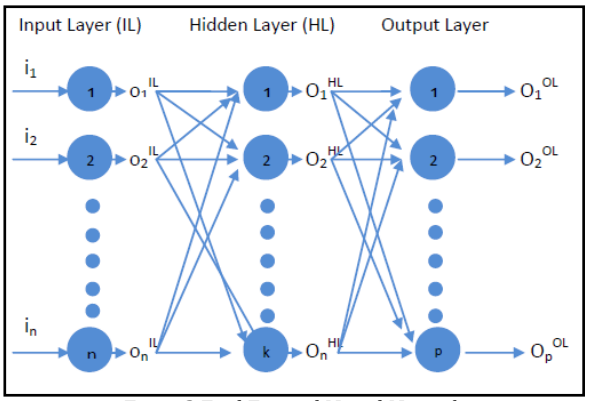

Figure 3 Feed Forward Neural Network

The output of input layer(IL) is given by-

$$O_j^{IL} = s(i_j)$$

While the output of hidden layer(HL) is represented by-

$$O_j^{HL} = s(b_i^{HL} + \sum_{j=1}^{n} w_{j,i}^{HL} O_j^{IL})$$

And the output of output layer is given by-

$$O_j^{OL} = s(b_i^{OL} + \sum_{j=1}^{n} w_{j,i}^{OL} O_j^{HL})$$

IV RESULTS

Proposed language identification system is tested for three different languages i.e. Hindi, English and Chhattisgarhi. 500 samples of each language have been collected and 78 features is extracted out from each language and prepared a feature database. A feed forward neural network is designed with 78 input neurons, 10 hidden neurons and 3 output neuron (one for each language). A network is trained by providing the feature vector from the database to input layer and also the target vector. From the result it is clear that the proposed language identification system is able to surpass the overall accuracy of 85%. We tested the result on hindi Vs Chhattisgarhi, Hindi Vs English and Chhattisgarhi Vs English. In all cases the accuracy is found to be more than 85%.

Table1 Hindi Vs English

	Hindi	English	Accuracy(%)
Hindi	45	05	90%
English	04	46	92%

Table2 Hindi Vs Chhattisgarhi

	Hindi	**Chhattisgarhi**	**Accuracy(%)**
Hindi	45	05	90%
Chhattisgarhi	06	44	88%

Table3 Chhattisgarhi Vs English

	Chhattisgarhi	English	Accuracy(%)
Chhattisgarhi	44	06	88%
English	04	46	92%

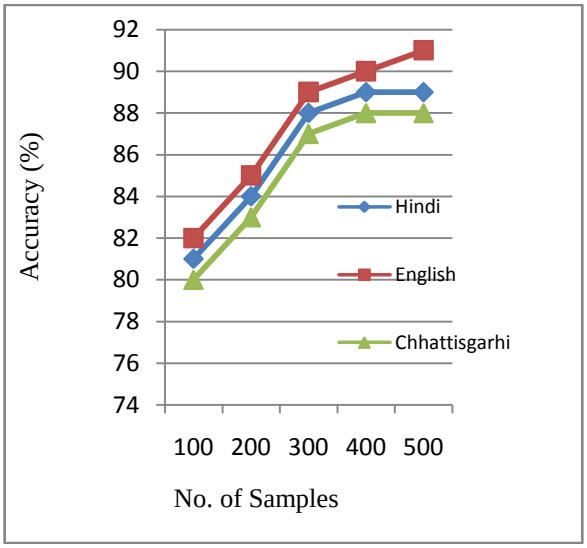

Figure 4 Over all Accuracy Graph for Languages

V. LIMITATION AND FUTURE SCOPE

Despite the good result obtained, there are many limitations in this approach. One of the problem of this method is noise free recording of spoken language which is difficult in real time implementation where some noise sources are present. Second problem is this that different person speaks the language in different manner creating a false representation of language. It is very difficult to determine the optimal number of hidden layer or neurons because it also varies with language to language. In future some more methods or neural network can be explored for better results.

REFERENCES-

[1] Ma, B., C. Guan, H. Li, and C.-H. Lee, "Multilingual Speech Recognition with LanguageIdentification," International Conference on Spoken Language Processing, 2002, pp. 505-508.
[2] Dai, P., U. Iurgel, and G. Rigoll, "A novel feature combination approach for spoken document classification with support vector machines," Multimedia Information RetrievalWorkshop, 2003, pp.1-5.
[3] Leonard, R.G., Doddington, G.R. Automatic Language Identification. Technical Report RADC-TR-74-200, Air Force Rome Air Development Center, August 1974.
[4] Leonard, R.G. Language Recognition Test and Evaluation.. Technical Report RADCTR-80-83, Air Force Rome Air Development Center, March 1980.
[5] House, A.S., Neuberg, E.P. Toward Automatic Identification of the Languages of anUtterance: Priliminary Methodological Considerations. Journal of the Acoustical Society of America, 62(3), pp. 708-713, 1977.
[6] Li, K.P., Edwards, T.J. Statistical Models for Automatic Language Identification. Proc.ICASSP'80, pp 884-887, April 1980.
[7] Cimarusti, D., Ives, R.B. Development of an Automatic Identification System of Spoken Languages: Phase 1. Proc. ICASSP'82, pp. 1661-1664, May 1982.
[8] Foil, J.T. Language Identification Using Noisy Speech, Proc. ICASSP'86, pp. 861-864,April 1986.
[9] Goodman, F.J., Martin, A.F., Wohlford, R.E. Improved Automatic Language Identificationin Noisy Speech. Proc. ICASSP'89, pp. 528-531, May 1989.
[10] Sugiyama, M. Automatic Language recognition using acoustic features. Proc. ICASSP'91, pp. 813-816, May 1991.
[11] Nakagawa, S., Ueda, Y., Seino, T. Speaker-independent, Text-independent Language Identification by HMM. Proc. ICSLP'92, pp. 1011-1014, October 1992.
[12] Muthusamy, Y.K. A Segmental Approach to Automatic Language Identification. PhD thesis, Oregon Graduate Institue of Science and Technology, October 1993.
[13] Yan, Y. Development of an Approach to Language Identification Based on Language dependent Phone Recognition. PhD thesis, Oregon Graduate Institue of Science and Technology, October 1995.
[14] Schultz, T., Rogina, I., Waibel, A. LVCSR-Based Language Identification. Proc. ICASSP'96, pp. 781-784, May 1996.
[15] Berkling, K., Reynolds, D., Zissman, M.A. Evaluation of Confidence Measures for Language Identification. Proc. Eurospeech'99, vol. 1, pp. 363-366, September 1999.