Identification of Indian Languages in Noisy Environments by using GMM

G. Sirisha
M.TECH, ECE Department
V R Siddhartha Engineering College
Kanuru, Vijayawada-520007, India.

P. Phani Kumar
Associate Professor, ECE Dept
V R Siddhartha Engineering College
Kanuru, Vijayawada-520007, India.

ABSTRACT
In this paper the identification of 27 Indian languages is performed using MFCC (Mel Frequency Cepstral Coefficients), GMM (Gaussian Mixture Models) and the accuracy of the language identified is checked in time varying noisy environments. The features are extracted using MFCC’s and from the features extracted the models are created using GMM. The models created in training phase are matched with GMM models in the testing phase. The percentage of all Indian languages is calculated by identifying matched samples. The noise that is collected from heavy traffic conditions which include Vehicle noise, horn noise, car noise, different pedestrians noise etc. is added to the speaker and tested. The efficiency is calculated when the noise is added and in order to increase the performance, Speech Enhancement technique Spectral Subtraction (SS) is used. The efficiency is also calculated after enhancement. Hence the efficiency is calculated for clean speech, noise speech, enhanced speech for 4, 8, 16, 32, 64, and 128 GMM models. The noisy speech and enhanced speech is evaluated for SNR = 0dB, 5dB and 10dB respectively.

For this IITKGP MLILSC (Indian Institute of Technology Kharagpur Multi Lingual Indian Language Speech Corpus) database which consists of 27 Indian languages is used. The experiments are carried on MATLAB environment which helps to obtain the results for clean speech, noisy speech, and enhanced speech.

Keywords: Mel Frequency Cepstral Coefficients (MFCC), Gaussian Mixture Model (GMM), Spectral Subtraction (SS), IITMLILSC Database, Enhancement.

I. INTRODUCTION
Language identification is the process of identifying the language from the given speech sample. Language identification plays an important role in real time applications like speech to speech translation systems, Interaction systems in driver assistance systems. The removal of noise also plays in real time applications like conversation in mobile, Bluetooth the person might not understand the other person’s conversation due to the noise associated in the surrounding environments. In this paper we are mainly concentrated for driver assistance i.e., while driver is driving the person goes through different types of background noise like Horn noise, vehicle horn noise, and different persons noises in real time environments. Hence this noise is added to clean speech and the performance is calculated. And the enhancement technique called Spectral Subtraction is performed to calculate the efficiency.

In this first the identification of 27 languages is performed and the accuracy is calculated .In this features are extracted using MFCC and GMM models are created in training phase and matched with features and models of testing phase. The efficiency is calculated from the correct samples assigned to each cluster with the specific mixture order for the complete speech database.

This paper is organised as follows: In Section II we have done brief literature survey on identification of languages using GMM. It has been observed that GMM is a very popular technique and is widely used for language classification along with various feature extraction techniques. In Section III we have given a description of the language corpus i.e. the database used for LID system. Section IV describes the language identification model consisting of detailed analysis of feature extraction using MFCC, classification using GMM and finally testing to recognise a particular language. In Section V we analyse the experimental results that are obtained and finally in Section VI we present our conclusions and scope for future work.
II. LITERATURE SURVEY

Language identification includes two phases namely training and testing. In the training phase features are extracted using Mel Frequency Cepstral coefficients and in the testing phase the features are matched and the performance of the language are determined. Many researchers worked in the field of language identification. Here we discuss the few recent studies based on language identification.

At first Automatic Language Identification: A tutorial [1] was proposed by Ambikairajah et al. which mainly concentrates on the progression of spoken language identification systems and current developments. Another Study was proposed by Torres-Carrasquillo et al. [2] which mainly concentrates on the two GMM based approaches to language identification that use shifted delta cepstra (SDC) feature vectors to achieve LID performance comparable to that of the best phone-based systems. Ming, Ji, et al [3] proposed the study based on Robust speaker Recognition, which deals with the elimination of the Environmental Noise while conversation with mobile or any internet devices with the combination of multi-condition model training and missing feature theory to model noise.

Based on GMM and robust language identification, Manchala et al. [4] proposed a paper. In this study Mel frequency Cepstral coefficients and formant frequencies are derived using the short time window speech signal and formant frequencies are extracted from LP analysis. And by using these features new feature vectors are generated by using cluster based computation and classification is done through GMM classifier and is carried on OGI database. Very few studies have been carried on Indian languages. In Indian context, Language Identification of Indian languages based on Gaussian mixture models Roy et al.[5] proposed a study which deals with only four Indian languages and the performance is calculated.

Another study by Sudhamay Maity et al.[6] which mainly concentrated on the prosodic features for discriminating the languages. When comes to Driver Assistance systems Hansen, John HL, et al. [7] proposed a study which concentrated on interaction between user and the in-vehicle system. It also concentrates with the removal of noise between the user and the interaction system.

III. LANGUAGE CORPUS

As this project is concentrated on the removal of noise in Indian languages, it is carried on a large database. The project is carried on IITKGP-MLILSC (Indian Institute of Technology Kharagpur Multi-Lingual Indian Language Speech Corpus Multi-lingual Language Speech Corpus) database which consists of 27 Indian languages namely Arunachali, Assamese, Bengali, Bhojpuri, Chhattisgarhi, Dogri, Gojri, Gujarati, Hindi, Indian English, Kannada, Kashmiri, Konkani, Manipuri, Mizo, Malayalam, Marathi, Nagamese, Nepali, Oriya, Punjabi, Rajasthani, Sanskrit, Sindhi Tamil, Telugu and Urdu. The database is prepared with the 20 speakers i.e., 10 male speakers and 10 female speakers. The database is collected from news bulletins, talk shows, interviews, live shows.

IV. PROPOSED WORK

The process of Language Identification includes two phases namely training and testing. Training includes the feature extraction, GMM modelling. Feature extraction means the 13 feature vectors are extracted using MFCC and from the features extracted the GMM models are created. In testing, the test data is converted into samples and from the samples the feature extraction is performed. GMM models created are tested with the language models at training phase. Hence the performance is calculated for correct and matched samples between training and testing phase and the accuracy is calculated. In this we use 20 speakers for each language and among 20 speakers how many samples are detected and to find the accuracy for whole GMM model is the main idea of the paper. Language identification can be done by two approaches: Speaker dependent system and Speaker Independent system. In the speaker dependent model, the system should be tested with the same speech samples from the same person who were used in training phase while in the case of speaker independent model the system is tested with unknown speech samples. In our case, we have used the speaker independent model as it gives better identification accuracy. The block diagram of the language identification for 27 Indian languages is shown in Fig 1.
The Block diagram of the language identification without adding noise is given above which includes the following steps. A Language Identification (LID) system has three major components database preparation, feature extraction and classification. The recognition performance also heavily depends on the performance of the feature extraction technique used. Here we use (MFCC) Mel frequency Cepstral coefficients for feature extraction. LID Model is divided into various phases consisting of feature extraction, backend classification using GMM and finally testing to recognize a particular language. The testing is done with GMM models 4, 8, 16, 32, 64, 128 and checks the performance of the language identified.

Language Identification After Adding The Noise

In this paper as we concentrate on robust language identification so we add different kinds of noise that a driver goes through that means the noise that includes horn noise, voices of different persons, vehicle noise etc. After adding the noise the block diagram of the language identification is given Fig 2. Here we use the spectral subtraction technique for enhancing the speech signal and to reduce the noise effect.

Fig 2: Block diagram of language identification of after adding the noise
V. EXPERIMENTAL RESULTS

We have given 20 speakers for 27 Indian languages. That means total 540 samples are taken. And the accuracy is calculated by the samples identified among 540 samples. The accuracy is calculated by using the formula Accuracy of the language Identified = (correct number of samples/total)*100.

where correct number of samples means the number of samples correctly matched, Total means the total number of samples i.e., 540 samples.

The experiments are carried under Matlab environment which helps to obtain the results easily and effectively. Coming to experimental results they can be classified into two approaches. They are:

1. Finding the accuracy and percentage of the language identified without adding noise.
2. Finding the accuracy and percentage of the language identified by adding the noise.
3. Finding the accuracy and percentage when the enhancement technique called spectral subtraction is applied.

<table>
<thead>
<tr>
<th>GMM Model</th>
<th>Percentage of language identification (Clean data)</th>
<th>Percentage obtained for Noisy data</th>
<th>Percentage obtained for Enhanced data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0dB</td>
<td>5dB</td>
<td>10dB</td>
</tr>
<tr>
<td>4</td>
<td>60.18</td>
<td>5.18</td>
<td>11.66</td>
</tr>
<tr>
<td>8</td>
<td>63.70</td>
<td>6.48</td>
<td>9.44</td>
</tr>
<tr>
<td>16</td>
<td>62.77</td>
<td>5.18</td>
<td>10.74</td>
</tr>
<tr>
<td>32</td>
<td>63.70</td>
<td>6.48</td>
<td>12.77</td>
</tr>
<tr>
<td>64</td>
<td>59.44</td>
<td>5.55</td>
<td>14.07</td>
</tr>
<tr>
<td>128</td>
<td>55.18</td>
<td>7.22</td>
<td>12.77</td>
</tr>
</tbody>
</table>

Table I

The performance is calculated with 0dB, 5dB, 10dB SNR values and the above results are obtained. The graph is obtained for the Clean data, Noisy data and Enhanced data for SNR=0dB. The software that we are using is Matlab. The obtained results of the clean data, noise data, enhanced data which is given in Fig 3.

Fig. 3 Comparative analysis of clean speech, noisy speech enhanced speech
Results Analysis

From the table I we can say that the performance of clean data is fallen when added with noise and after enhancing by using Spectral Subtraction the accuracy and the performance is increased. The graph also shows the clean speech spectrum, noisy speech, and enhanced speech. It clearly states that the performance is increased from the noisy speech by using enhanced technique called Spectral Subtraction. In the graph, the variation is clearly observed at 0.5,1 ...etc durations.

V. DISCUSSIONS AND CONCLUSIONS

To conclude it is clear that at 8 and 32 Gaussian mixture model the performance is high. The performance can be further increased by changing the number of speakers individually and check the performance individually. The performance of the robust language identification is improved further by using other different and new enhancement techniques like Voice activity Detection, Multi-Microphone Arrays and multi channel speech enhancement techniques also. Here in this paper we are calculating the only overall percentage of the language GMM model. The performance of the individual languages can also be checked in the future scope i.e., the number of samples matched out of number of samples trained. This application software will be embedded in onboard unit and it can be used as driver assistance systems and it is used for the driver modelling.

REFERENCES


