

# Linguistic Differentiation and Identification Using Speech Processing

Atif Imteyaz

School of Electrical Engineering  
MIT Academy of Engineering  
Alandi(D), Pune, INDIA

Tejas Joshi

School of Electrical Engineering  
MIT Academy of Engineering  
Alandi(D), Pune, INDIA

Tina Kalambhe

School of Electrical Engineering  
MIT Academy of Engineering  
Alandi(D), Pune, INDIA

Uday Mithapelli

Assistant Professor  
School of Electrical Engineering  
MIT Academy of Engineering  
Alandi(D), Pune, INDIA

**Abstract-** The speech signal conduct information about the identity of the speaker. The area of speaker information cares with extract the identity of the person express the utterance. The utility of automatically identifying a speaker is predicated solely on vocal characteristic. This emphasizes on text dependent talker identification, which deals with detecting a specific speaker from a known region. System identifies the user by comparing the code book of speech utterance with those of the stored within the database and lists, which contain the foremost likely speakers, could have given that speech utterance. The speech signal is recorded for number of speakers further the features are extracted. Feature extraction is completed by means of LPC coefficients, calculating AMDF, and DFT. The neural network is train by applying these features as input parameters. The train network corresponds to the output, the input is that the extracted features of the speaker to be identified. The network does the weight adapting and the best match is found to identify the speaker of the region.

**Keywords**— short utterance feature representation, spoken language identification, dataset.

## I. INTRODUCTION

Language identification and multilingual speech recognition are key to the event of spoken dialogue systems which will function in multilingual environments. Several approaches to language identification have proposed within the literature. System supported Gaussian mixture models classify the speech features, more sophisticated class of language identification systems makes use of continuous speech recognition algorithms. One of the ways is described in our paper which uses speech processing as a way of linguistics identification and differentiation. In speech processing, speech is processed on a frame by-frame basis usually only with the priority that the frame is either speech or silence. The usable speech frames are often define as frames of speech that contain higher information content compared to unusable frames with reference to a selected application. we've been

investigating a talker identification system to recognized usable speech frames. We then determine how for identifying those frames as usable employing a special approach. speech Identification (LiD) is that the process of recognition of the language spoken in an utterance. Automatic language identification is that the matter of identifying the language being spoken from a sample of speech by a speaker. As with speech recognition, humans are the foremost perfect language Identification systems within the world today. Within seconds of hearing speech, people are able to determine whether it's language they know. If it is a language with which they are not familiar, they often can make instinctive judgments on its similarity to a language they know. Any utterance is nothing but a speech or audio signal. Speech processing is that the research of speech signals and thus the processing methods of these signals. The signals are usually clarified during a digital representation, so speech processing are often considered a special case of digital signal processing.

## II. LITERATURE REVIEW

After studying various Research Paper, we learned about Speech Recognition, Speech Verification and Speech Identification and different approaches to Speech recognition. Dialect identification is an important research topic in speech recognition. Speaker Recognition accept which of the population of subjects spoke a given utterance. Speaker verification confirm that a given speaker is one who he claims to be. System encourages the user who claims to be the speaker to supply ID. System verifies user by comparing codebook of given speech utterance thereupon given by user. If it matches the set threshold then the identity assert of the user is accepted otherwise rejected. Speaker identification detects a specific speaker from a known population. System associates the user by comparing the codebook of speech utterance with those of the stored.

within the database and lists, which contain the foremost likely speakers, compared with speaker recognition and language identification (LiD) Dialect can be recognized by a speaker's phones, pronunciation, and traits such as tonality, loudness, and nasality. This paper emphasizes on text dependent talker identification, which deals with

detecting a specific speaker from a known population. The system prompts the user to supply speech utterance. System recognized the user by comparing the codebook of speech utterance with those of the stored within the database and lists, which contain the foremost likely speakers, could have given that speech utterance. The speech signal is recorded for N speakers further the attribute is extracted. For LID and dialect identification, i-vector is regarded as the state-of-the-art for general tasks. Based on the joint factor analysis technique (JFA), it was proposed that speaker and session differences can be characterized by a single subspace.

Ms. Vimala.C and Dr. V. Radha proposed speaker independent isolated speech recognition system for Tamil language. Feature extraction, acoustic model, pronunciation dictionary and language model were implemented using HMM which produced 88% of accuracy in 2500 words. Suma Swamy et al. introduce an efficient speech recognition system which was experimented with Mel Frequency Cestrum Coefficients (MFCC), Vector Quantization (VQ), HMM which recognize the speech by 98% accuracy. The database consists of five words spoken by 4 speakers at ten times. Maya Money Kumar, et al. developed Malayalam word identification for speech recognition system. The proposed work was done with affricate-based segmentation using HMM on MFCC for feature extraction

### III. PROPOSED METHODOLOGY

**Lemmatization:** It assume reducing the various inflected forms of a word into a single form for easy research.

**Morphological segmentation:** It elaborate dividing words into individual units called morphemes.

**Word segmentation:** It involved dividing a huge piece of continuous text into distinct units.

**Part-of-speech tagging:** It involves recognize the part of speech for every word.

**Parsing:** It convoluted undertaking grammatical analysis for the provide sentence.

**Sentence breaking:** It involves set sentence boundaries on a large piece of text.

**Stemming:** It intricate cutting the inflected words to their root form.

- 1)
1. pre-processing
2. Hidden Markov model (HMM)
3. Neural Networks Artificial neural networks (ANN)
4. Natural Language Processing

These three stages together contribute to the language identification. And use of machine

learning inherently calls for two phases of operation,

1. training.
2. testing.

The system is first instructing with the available dataset and then tested with samples.

**Pre-processing:** Pre-processing is the modulate stage of the system. In the pre-processing stage various methods are acquires to bring all the input data at the same configuration of the concerned fields or attributes. The basic pre-processing involves background noisereduction and resampling.

The pre-processing performed has two steps.

- A. re-sampling and
- B. file format handling.

The Gaussian mixture model (GMM) is commonly used for determining how well each HMM state fits a frame of the acoustic input, i.e. the probability, and with enough components, they can model probability distributions to any level of accuracy. The accuracy of a GMM-HMM system can be improved further with fine-tuning after it has been trained.

We calculated the word error rate (WER) and accuracy according to the equations.

$$WER = (I + D + S) / N$$

where I stand for inserted, D stands for deleted, and S stands for substituted.

**Feature extraction:** Transforming the input features into the set of features is called feature extraction. Feature extraction is a very crucial stage in language identification system. If the features extracted are consciously chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation rather of the full-size input. When performing analysis of complex data one of the major problem stems from the number of variables intricate.

Analysis with a huge number of variables normally need a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for techniques of build combinations of the variables to get around these problems while still describing the data with sufficient accuracy. With respect to language identification the first

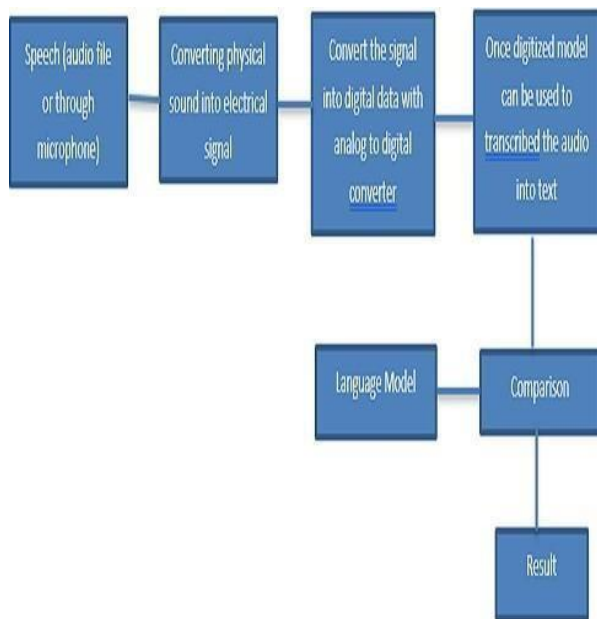
function is to identify features which may provide us with information relevant to the task at hand. The audio features can be assorted as high level and low level. Low level features are the ones which can be directly obtained from the speech samples without any extra operations.

High level features are the ones that are taken from the audio by performing mathematical operations like transformations on them. The feature extraction is not a one-step extraction process but intricate many sequential phases.

The feature extraction stage of the Lid is represented here. The data coming in after pre-processing undertaken the following steps.

1. windowing
2. Discrete Fourier Transformation
3. Mel filter bank
4. Discrete Cosine Transform
5. Mean MFCC.

#### IV DESIGN OF THE SYSTEM



#### V. MODULE-WISE DESCRIPTIONS

Pre-processing is the recording of speech with a sampling frequency of, for example, 16 kHz and, according to The Shannon Theorem, a bandwidth limited signal can be reconstructed if the sampling frequency is more than dual the maximum frequency meaning that frequencies up to almost 8 kHz are constituted correctly.

In development of an ASR system, pre-processing is examining the first phase of other phases in speech recognition to differentiate the voiced or unvoiced signal and create feature vectors. pre-processing calibrates or modify the speech signal,  $x(n)$ , so that it will be more acceptable for feature extraction analysis. The major element to consider when it

comes to speech signal processing is to check the speech,  $x(n)$  if is corrupted by some background or ambient noise,  $d(n)$ , for example as additive disturbance.

$$x(n)=s(n)+d(n)$$

Where  $s(n)$  is the clean speech signal. In noise reduction, there are different techniques that can be adopted to perform the task on a noisy speech signal. However, to evolve perfect speech recognition system, the two frequently used methods of noise reduction algorithms in speech recognition system is spectral subtraction and adaptive noise cancellation ([D+00]).

Hidden Markov model (HMM) is defined as” a doubly stochastic process with an underlying stochastic process that is not visible (it is hidden), but can only be observed through another set of stochastic processes that make the sequence of observed symbols”.

Neural Networks Artificial neural networks (ANN) are, as the name implies, inspired by the sophisticated functionality of the human brain where neurons process information in parallel. ANN consists of an layer of input nodes, then one hidden layer of nodes and finally an layer of output nodes.

Instead of the N-gram language model, we can construct neural language models and feed them into a speech recognition system to restore things that were produced by a first path speech recognition system.

focus into the accent model, we can figure out how to do pronunciation for a new sequence of characters that we’ve never seen before using a neural network.

For acoustical model, we can build deep neural network to get much better categorization accuracy results of the feature for the current frame.

Interestingly sufficient, even the speech pre-processing steps were established to be essential with convolutional neural networks on raw speech signals.

Natural Language Processing (NLP) is a branch of artificial intelligence that distribute with the interaction between computers and humans using the natural language. Most NLP techniques rely on machine learning to derive sense from human languages.

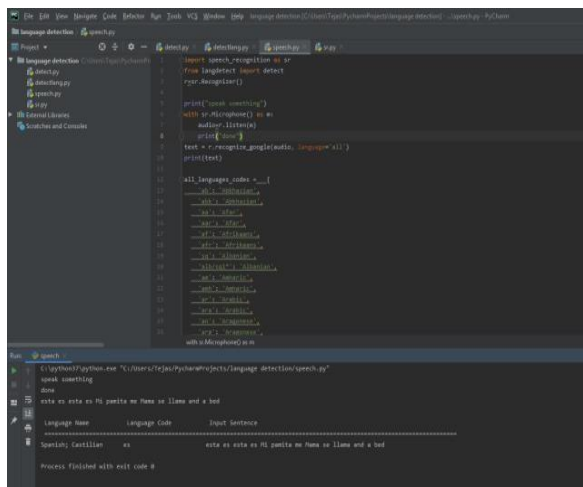
Natural Language Processing (NLP) speech to text is a extreme application of Deep Learning which allows the machines to understand human language and read it with a motive to act and react, as usual, humans do. The basic idea behind NLP is to nourish the human language as in the

form of data for intelligent systems to consider and then utilize in various domains. Natural Language processing has made it possible to mimic another main human trait

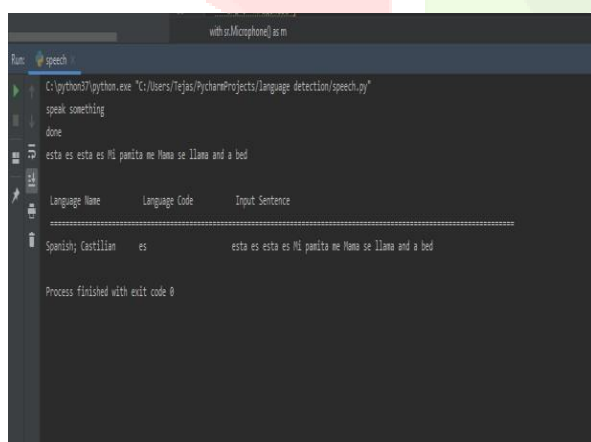
i.e., conception of language and has made it possible to bring about all the transformational technologies 1.

some examples of such are Alexa and Siri on a more trending scale and autonomous call Centre agents on a more operational scale. NLP is usually deployed for two of the primary function namely Speech Recognition and Language Translation. Google translator is one of the most ordinary examples of Natural Language Processing. Using the deep learning algorithms and in discrete the Neural Networks, the NLP can do a lot with the unstructured text data by finding patterns of sentiments, major phrases used for certain situations, and specific text slates within a block of text.

## VI. IMPLEMENTATION



## VII. RESULT



## VIII. CONCLUSION AND FUTURE SCOPE

In this project, we learned about several things such as working of Speech Recognition system and its applications. The results obtain appear that the deviation in the results for the same speaker speaking at different instances is negligible. The software works fine for identifying speaker from number of different dialects. The vocabulary is limited Number of users are limited Also, for the project work, we had to go to various data and references for the completion of our project.

## REFERENCES

1. K. M. Berkling, T. Arai and E. Barnard, "Analysis of phoneme-based features for language identification", in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 94, Adelaide, Australia, April 1994
2. J. Hieronymous and S. Kadambe, "Spoken Language Identification Using Large Vocabulary Speech Recognition", in Proceedings of the 1996 International Conference on Spoken Language Processing (ICSLP 96), Philadelphia, USA, 1996.
3. K. M. Berkling and E. Barnard, "Language Identification of Six Languages Based on a Common Set of Broad Phonemes", in Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP 94), Yokohama, Japan, September 1994.
4. K. M. Berkling and E. Barnard, "Theoretical Error Prediction for a Language Identification System using Optimal Phoneme Clustering", in Proceedings 4rd European Conference on Speech Communication and Technology (Eurospeech 95), Madrid, Spain, September 1995
5. Rodriguez, J. E. F. F. R. E. Y. J., Lim, J., & Singer, E. (1987, April). Adaptive noise reduction in aircraft communication systems. In ICASSP'87. IEEE International Conference on Acoustics, Speech, and Signal Processing (Vol. 12, pp. 169-172). IEEE.
6. Wang, H., Lu, Z., Zhao, H., & Feng, H. (2015, September). Application of Parallel Computing in Robust Optimization Design Using MATLAB. In 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC) (pp. 1228- 1231). IEEE.
7. 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)
8. F. Biadys, J. Hirschberg, and M. Collins, "Dialect recognition using a phone-gmm- supervector-based svm kernel," in Proceedings of Interspeech, 2010,
9. .N. F. Chen, W. Shen, J. P. Campbell, and P. A. Torres- Carrasquillo, "Informative dialect recognition using context- dependent pronunciation modeling," in Proceedings of ICASSP, May 2011.
10. R. Tong, B. Ma, H. Li, and E. S. Chng, "Target-aware lattice rescoring for dialect recognition," in Proceedings of Interspeech, 2011