Isolated Speech Recognition System

P. N. Bendkoli[1] Prof. P. B. Borole[2]
[1,2]Department of Electrical Engineering
[1,2]V.J.T.I. Mumbai, India

Abstract— The main objective of this project is to design and implement a real time speech recognition system. “Real time speech to text” can be defined as accurate conversion of words that represents uttered word instantly after speaking. The system is design in such a way that when the user speaks in front of microphone then incoming speech samples are compared with the stored samples in the system memory and if they matched then the respective word will be shown on the computer screen. For the system implementation we are using the MATLAB. This system is based on the Mel Frequency Cepstral Coefficients (MFCC) and a distance-based matching algorithm.

Key words: MATLAB, MFCC

I. Introduction

Speech recognition has of late become a practical technology. Speech recognition is used in real-world human language applications. From the conversation or speech, it converts an acoustic signal that is captured by a microphone or a telephone, to a set of words. Speech recognition algorithms can be broadly divided into speaker dependent and speaker independent. Speaker dependent system focuses on developing a system to recognize unique voiceprint of individuals. Speaker independent system involves identifying the word uttered by the speaker. It can be further classified into isolated word detection and continuous speech recognition. This paper represents the isolated speech recognition system. . Inputs for this system are the isolated words separated by pauses. When the user utters something, it is sent to the speech engine to be processed then converted into digital domain. The digitalized speech samples are processed to extract features using MFCC algorithm. Once the desired number of features is obtained, they can be sent through feature matching stage where DTW is used for comparison between saved templates and recorded speech this entire system is implemented in MATLAB environment.

II. Literature Survey

Speech Recognition Approaches- Automatic speech recognition system is used to transform or produce a sequence of message from a speech signal. This process is called decoding. Speech signal is decoded and converted converted into writing or commands to be processed. In general, there are three classical approaches as follows: Acoustic-phonetic approach, Artificial intelligence approach, Pattern recognition approach. The acoustic-phonetic approach is based on the theory of acoustic phonetics. The theory proposes that there exist finite, distinct phonetic units in spoken language. The phonetic units are characterized by a set of properties that are embedded in the speech signal or its spectrum. The artificial intelligence approach is a compound approach that utilizes the ideas of the first two approaches

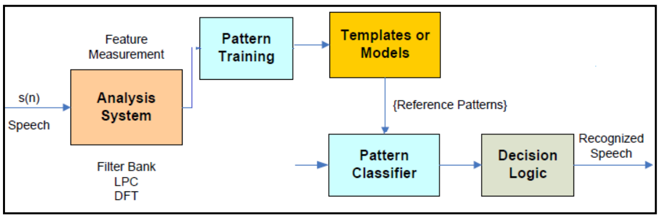

Fig. 1: Pattern Recognition Approach

In pattern-recognition approach, the speech patterns are used directly without explicit feature determination and segmentation. There are two main steps: Training of speech patterns and recognition of patterns via pattern comparison. Speech knowledge is supplied into the system via the training procedure. Most of the current and modern ASR systems are based on the principles of statistical pattern recognition. As shown in fig 1 firstly, features are extracted from the input signal and represented into a form of features. There are number of spectral analysis techniques, such as filter DFT, LPC and MFCC analysis. Secondly, in pattern training one or more test patterns corresponding to speech sounds of the same class are used to create a pattern representative of the features of that class. The resulting pattern, generally called reference pattern, can be template, derived from some type of averaging technique, or it can be a model that characterizes the statistics of the features of the reference pattern.

This is followed by pattern classification process. Here, the unknown speech input is compared in pattern training, by measuring the similarity between the train and test pattern. Lastly, decision logic is applied to decide which reference best matches the unknown test pattern.

III. System Design

A. System Components

The prepared system if visualized as a block diagram will have the following components: Sound Recording component, feature extraction component, speech recognition component

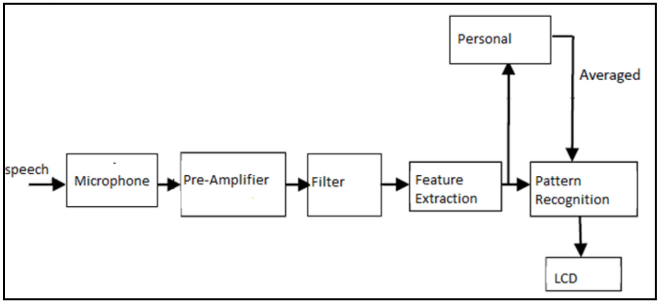

Fig. 2: Block Diagram of Speech Recognition System

1) Sound Recording

The component is responsible for taking input from microphone and identifying the presence of words. Word detection is done using energy and zero crossing rate of the signal. The output of this component can be a wave file or a direct feed for the feature extractor

2) Feature Extraction component

The component generated feature vectors for the sound signals given to it. It generates Mel Frequency Cepstrum Coefficients and Normalized energy as the features that should be used to uniquely identify the given sound signal

3) Recognition Component

This is a Continuous, Multi-dimensional Hidden Markov Model based component. It is the most important component of the system and is responsible for finding the best match in the knowledge base, for the incoming feature vectors.

Once the training is done, the basic flow can be summarized as the sound input is taken from the sound recorder and is feed to the feature extraction module. The feature extraction module generates feature vectors out of it which are then forwarded to the recognition component.

The recognition component with the help of the knowledge model and comes up with the result. During the training the above flow differs after generation of feature vector. Here the system takes the output of the feature extraction module and feeds it to the recognition system for modifying the knowledge base.

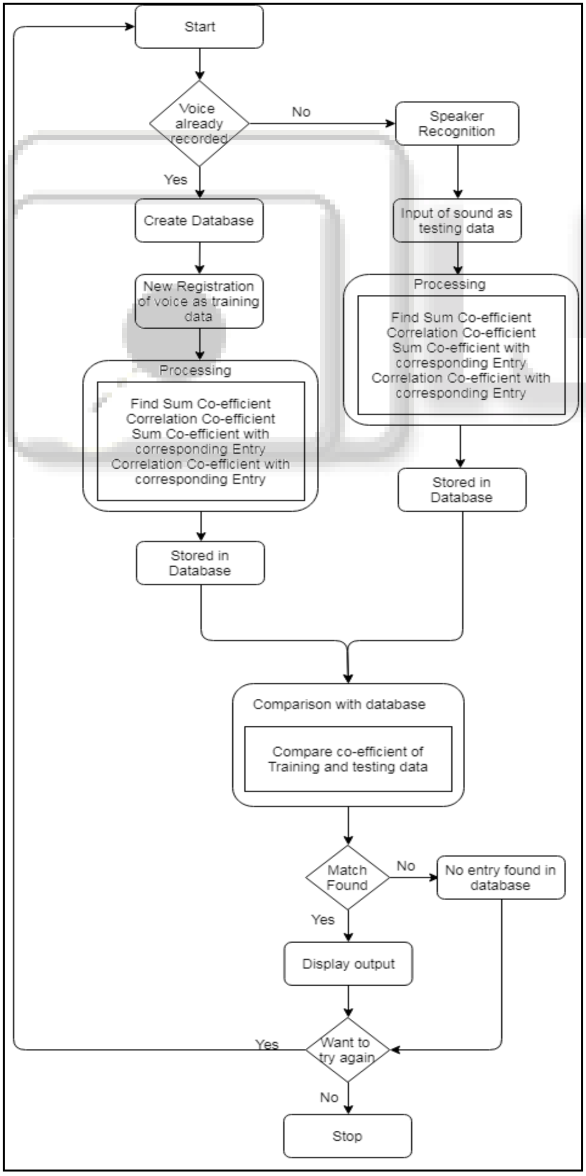
Start
Voice already recorded
No
Yes
Speaker Recognition
Create Database
Input of sound as testing data
New Registration of voice as training data
Processing
Find Sum Co-efficient
Correlation Co-efficient
Sum Co-efficient with corresponding Entry
Correlation Co-efficient with corresponding Entry
Processing
Find Sum Co-efficient
Correlation Co-efficient
Sum Co-efficient with corresponding Entry
Correlation Co-efficient with corresponding Entry
Stored in Database
Stored in Database
Comparison with database
Compare co-efficient of Training and testing data
Match Found
No
No entry found in database
Yes
Display output
Yes
Want to try again
No
Stop

Fig. 3: Flowchart of the Working System

B. Woking of the System

As this system is fully implemented in MATLAB therefore the flow of the working of the system is as follows.

1) Creating the Database

Database is created by recording the speech samples into the system though the training process. In this speech signal is given through the microphone of the computer system. Then the speech signal is then digitalized for the further processing.

2) Extracting Features of the Recorded Data

Useful information is extracted from the recorded data by calculating the sum and correlation coefficients of the digital samples of the analog speech waveform. This analog to digital conversion of the speech sample is also done in the MATLAB. These calculated coefficients are then stored in the database.

3) Testing the Real Time Speech Signal

In this, steps involved in the second point are repeated and sum and correlation coefficient of the real time signal are calculated and stored in the database.

4) Comparison of Both the Databases

In this process comparison of the sum and correlation coefficient of the stored speech samples and the real speech samples of testing purpose is carried out and if the match found then the respective word corresponding to that coefficient is displayed on the computer screen.

C. Building Process

In the process of completing our project, we have to go through certain processes. Those processes are shown below:

1) Recording audio & converting it into .wav format
2) Processing that .wav file
3) Storing it in a file
4) Making software to compare the audio with other audio files with inserted voice and recognize it
5) Making a program to show the voice files in text format.

IV. RESULTS, CONCLUSION & FUTURE SCOPE

In the results, the small dictionary of the words is recognized Dictionary words are one, two, three, four, and five. Recognition of these words has been successfully done with the existing system,

The system performance is sensitive to the amount of training data available for creating speech reference patterns. The more training, the higher the performance of the system. The reference patterns are sensitive to the speaking environment and transmission characteristics of the medium used to create the speech, this is because the speech spectral characteristics are affected by transmission and background noise.

In future the accuracy of the system can be increased by using more advanced techniques available feature extraction and better results can be obtain regardless of the background noise.

REFERENCES

[1] Salih M. ,Al-Qaraawi, Sarah Mahmood “Wavelet Transform Based Features Vector Extraction in Isolated Words Speech Recognition System” 2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)

[2] K. Sung-Nam, H. In-Chul “A VLSl CHIP FOR ISOLATED SPEECH RECOGNITION SYSTEM” IEEE Transactions on Consumer Electronics, AUGUST 1996, Vol. 42, No. 3

[3] Mohsen S. Hossein Marvi,”Optimal MFCC Feature Extraction By Differntial Evoluation Algorith for Speaker Recognition” 2017 3rd Iranian Conference on Signal Processing and Intelligent Systems

[4] Nuzhat Atiqua Nafis, Md. Safaet Hossain, “Speech to Text Conversion in Real-time”, International Journal of Innovation and Scientific Research

[5] Chadawan Ittichaichareon, Siwat Suksri and Thaweesak Yingthawornsuk,”Speech Recognition Using MFCC”, International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand)

[6] Ms. Vrinda1 ,Mr. Chander Shekhar,”Speech Recognition System For English Language” International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2013

[7] Su Myat Mon, Hla Myo Tun, “Speech to Text (STT)Conversion Using Hidden Markov Mode(HMM)l” International Journal of Scientific & Technology Research Volume 4, issue 06, june 2015