International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 03 | Mar -2020 ISSN: 2582-3930

# DEVELOPMENT OF MARATHI SPEECH BASED SYSTEM FOR HUMAN ROBOT INTERACTION

1st Author :Ms.NikitaSakpal
Department of Computer engg
Siddhant college of Engg,sudumbare,
Pune,India.

2nd Author :Prof.Manisha Darak
(AssistantProfessor)
Department of Computerengg,
Siddhant college of Engg ,Pune

***Abstract*—Now a days various smart applications anddevices are present having English voice as input butEnglish is not a local language in Maharashtra. TheEnglish illiterates are not efficiently communicates with asystem because of language barrier. Our system overcomesthis problem and found solutionfor Marathi speakingpeople all over world. Proposed system is composed ofhardware components like HM2007 and Raspberry Piwhich are essential for voice recognition and synthesis ofMarathi language. With the help of the motors and**
**Ultrasonic sensor the system will detect obstacle andcommunicate with the user. We have also used the opensource software like Festival and Festvoxfor generating thevoice from Marathi text. We will also analyse the voicesample efficiency of user given input. So this system willhelp the Marathi peoples for executing their task inMarathi language.**

***Keywords-raspberry pi, hm2007, motor, l239d, voice recognition, text-to-speech, marathi speech synthesis, python***

## I. Introduction

Since ten decades, research in Automatic Speech Recognition System has made significant development such as Microsoft Translator, Google voice search, Apple's Siri etc. This system is now used with great interest in both industry and academia. [1].
In voice recognition and communication system, speech recognition is actively considered, and it has tremendous progress due to digital signal processing. Any speech recognition system consists of three fundamental steps- pre-processing, feature extraction and classification.
Surrounding and noise makes speech recognition system complicated as it affects a lot on performance and accuracy of the Automatic Speech Recognition
System. Speech Recognition accuracy is used for performance measurement of speech recognition system [2].Human uses a speech as one of the effective medium for communication. Speech is a digital signal made up of various components like time, amplitude and frequency. Because of this, in different frequency bands, transitions occur at different time. In simple word, we can defineAutomatic Speech Recognition (ASR) as a system which extracts, recognize and translate properties of speech using computer device. The purpose of the developing ASR system is to enable human real time interaction with a machine in natural language regardless of speech accent, speech noise, environment and size of vocabulary. ASR systems can be developed for recognizing isolated words, connected word and continuous speech. ASR systems has various industrial and academia applications such as telecommunication, for impaired hearing people, learning language for children's etc. [3]. Speech is a multi-component signal with varying time, frequency and amplitude. Due to this variability, transitions may occur at different times in different frequency bands. The technology developed for extraction, recognition, and translation of the speech characteristics by using the smart computerized device is called an Automatic Speech Recognition (ASR). The main purpose of ASR is to develop a technology which helps human to interact with the machine in our natural language in real time environment regardless of vocabulary size, noise, speech characteristics or accent[51, 52]. ASR can be implemented for isolated word recognition, connected word recognition, and continuous speech recognition. This paper is divided into five sections. First section discusses introduction to Speech Recognition System, second section discusses overview of Automatic Speech Recognition System, in third section, literature review, fourth section discusses about ASR systems for Indian languages and paper is concluded in fifth section.

Marathi is one of the most widely spoken language of the world (It is ranked between four and seven based on the number of speakers), with nearly 100 million native speakers. However, this is one of the most resource language which lack speech application to solve above problem we have proposed model which will beneficial for the English illiterates blind people in Maharashtra. The proposed system can be modified as per the user. Many researchers have been done experimentation in the research for getting expressive speech synthesis. Proposed model will sort out the problems of English illiterates and also for the blind Marathi speaking people in all over the world. The main idea behind the project is to make use of available technology for the generation of Marathi speech using computer software like Festival, Festvox. We have also used the hardware components like Raspberry pi development board for the processing the given input voice commands. In this work the motors will play

## II. LITERATURESURVEY

| Sr. No. | Title | Authors name | Year | Review |
|---|---|---|---|---|
| 1 | Minimum Prediction Residula Applied to Speech Recognition | Itakura et. al. | 1975 | Proposed the idea of using LPC which is a data compressing method to speech recognition system. |
| 2 | Speech recognition using LPC analysis | Ostrander et. al. | 1982 | Developed English digit speech recognition system using MFCC as extraction method and HMM model. |
| 3 | Speech recognition using the probabilistic neural network | Low et. at. | 1998 | Describes about implementation of speech recognition system on a mobile robot for controlling movement of the robot. The methods used for speech recognition system are Linear Predictive Coding (LPC) and Artificial Neural Network (ANN). Back propagation method is used to train the ANN. |
| 4 | Voice recognition algorithms using mel-frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques | Muda et al. | 2010 | Has talked about two speech acknowledgment model which are vital in speech recognition system they are MFCC and DTW. The non-parametric strategy for demonstrating the human sound-related observation framework, Mel Frequency Cepstral Coefficients (MFCCs) is use as extraction strategies. The nonlinear grouping arrangement known as Dynamic Time Warping (DTW) presented by Sakoe Chiba has been utilized as components coordinating procedures. |

| | | | | |
|---|---|---|---|---|
| 5 | Techniques for feature extraction in speech recognition system: A comparative study | Shrawankar | 2013 | Has discussed some features extraction techniques and their advantages and disadvantages. Some hybrid feature extraction methods discussed in the paper are RASTA-PLP, MFCC-LPC and HFCC-E. MFCC feature requires less computational time and it is robust for noise free environment. |

not to generalize a mapping from inputto output.(c) Semi-Supervised learning combines labeled as well as unlabeled examples togenerate a suitable function or classifier.

## III. PROPOSED METHODOLOGY

*A .Architecture:*

1. Machine Learning

Machine learning deals with the development of algorithms, which helps machine to learnby inductive inference which is based on the observations data which representincomplete information of thestatistical phenomenon. Classification, also known aspattern recognition, is a very important task inMachine Learning, with this machine“learn” to recognize complexes pattern automatically, itdistinguishes between exemplarswhich are based on their different patterns, and makes intelligentdecision. A patternclassification task has three modules namely data representation (featureextraction)module, feature selection or reduction module, and classification module. The firstmodule aims to find invariant features which describes the difference in the classes. Thesecond module of feature selection and feature reduction is used to decrease down thedimensionality of the feature vectors for classification. The classification module findsthe actual mapping between the pattern and label based on features. The objective of thepresent work is to investigate the machine learning method in the application of automaticrecognition of emotional state from human speech.
Different Machine Learning Algorithms are based on the input available at the time oftraining:
(a) Supervised learning algorithms are trained on labelled examples, i.e., input wherethe desired output is known. The supervised algorithm generalizes functions ormappings from input to outputs, thus it can be used to hypothetically generate anoutput for previous unseen inputs.(b) Unsupervised learning algorithms operate on unlabelled example, i.e., inputswhere the desired outputs are unknown. Here the objective is to discover structurein the data (e.g. through a cluster analysis),

*B. Algorithm*
There are two basic steps in any supervised learning system, testing and training. For any speech recognition system basic steps in training and testing are almost common and step in this are pre-processing, feature extraction and classification.

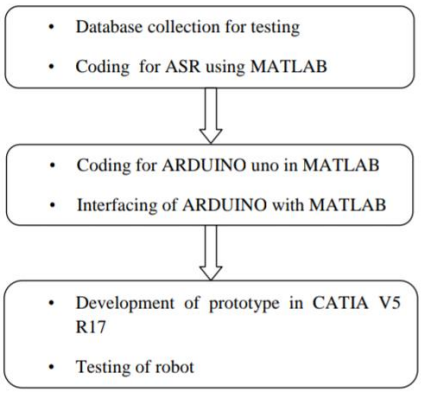



Figure 3.1: General outline of the work.

The general overview of the development of speech controlled wheel chair robot is
described in the figure 3.1. At first, speech samples of English commands (right turn, left
turn, start and stop) were collected using android mobile phone with 8 kHz sampling
frequency. Virtual testing and training speech recognition environment was developed in
MATLAB 13b version.
In this proposed model there are two phases training and testing, in the training phase,
training speech samples collected from the volunteers are processed and stored in the
database. And in testing phase unrecognized voice input is processed and recognized
using KNN and PNN classifier. Training and testing phases are shown in the figure

3.2.First the speech signal is pre-processed using first order FIR low pass filter and then framing and windowing is applied on speech signal, detail process is explained in latter part of the chapter. Voiced and unvoiced part is separated using prosodic features (Zero crossing rate, short term energy), then derived features (MFCC, LPC) are extracted from the signal. Features are quantized and the values are stored in database.

In testing all the above mentioned steps are followed and the quantized data is compared with the database using KNN/PNN classifier for recognition. This information of recognized voice is given to Arduino board using MATLAB Arduino interface. Arduino gives signal to L293D motor driver controller according to the programming done for different directions

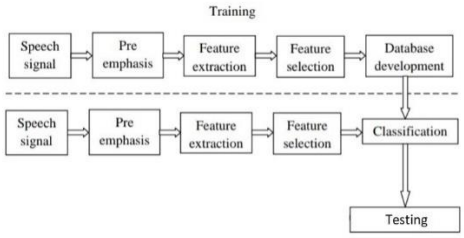



Figure 3.2: Speech Recognition System.

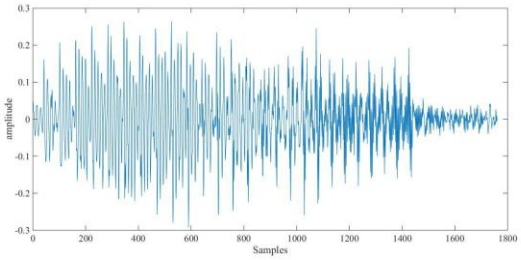
In the proposed work, we have used four possible combinations of isolated word recognition algorithm. Table 1 contains the details of the four combinations. In combination 1 speech input from the sample file is pre-processed after that LPC feature are extracted then features are classified using KNN classifier. In combination 2 all steps are same as in combination 1, only in step 3 12 MFCC features are extracted. In combination 3 all steps are same as combination 1 except step 5, in this PNN classifier is used instead of KNN. In combination 4all steps are same as combination 3 except step 3, in this MFCC is used instead of LPC. These four combinations are tested for speech samples from database.

| Combi-nation | Step 1 Input speech signal | Step 2 Pre-processing | Step 3 Feature extraction | Step 4 Feature selection | Step 5 Classification |
|---|---|---|---|---|---|
| 1 | Audio sample file | Low pass filter (FIR) | Linear prediction coding(LPC) | Split vector quantization | KNN classifier |
| 2 | Audio sample file | Low pass filter (FIR) | Mel-frequency cepstrum coefficient(MFCC) | Split vector quantization | KNN classifier |
| 3 | Audio sample file | Low pass filter (FIR) | Linear prediction coding(LPC) | Split vector quantization | PNN classifier |
| 4 | Audio sample file | Low pass filter (FIR) | Mel-frequency cepstrum coefficient(MFCC) | Split vector quantization | PNN classifier |

Table 3.1 various combination of feature for isolated word recognition system.

## IV. RESULT AND DISCUSSION

Automatic Speech Recognition (ASR) system is becoming the most important interface technique in the area of Robotics for the Socially Assistive Robots (SAR). Incorporation of ASR in wheel chair will increase comfort and simplify the controls. So, in the system speech signal is given through microphone to the MATLAB based program, it is executed and following outputs are obtained. For the command *"left"* output of digital FIR low pass filter is as shown in figure 4.1.




The signal is then segmented using zero crossing rate and short-term energy the figure 4.1 shows the segmented and original signals.

14 MFCC and 12 LPC feature vector as collected from the speech samples for feature classification. These features are quantized using split vector quantization to compress the data and efficiency are increased.

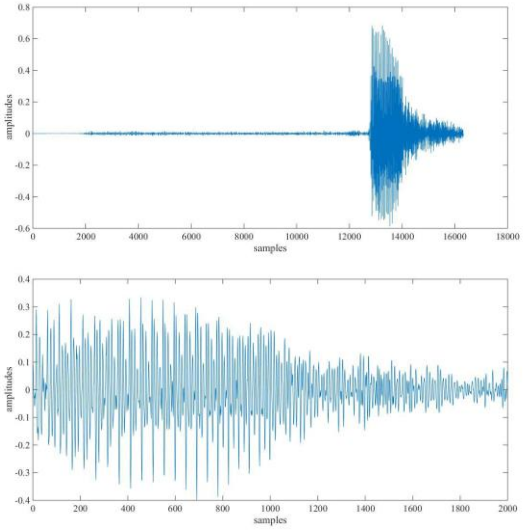



Figure 4.2: (a) Original signal and (b) voiced signal.

Table 4.1: Percentage accuracy of different combinations of algorithms.

| Sl. No. | Combinations | Efficiency (%) |
|---|---|---|
| 1 | Combination 1 | 72.5 |
| 2 | Combination 2 | 78.75 |
| 3 | Combination 3 | 76.25 |

4 Combination 4 83.75

In the above-mentioned table 4.1 accuracies of different proposed combinations are listed, the readings show that the combination 4 with MFCC feature extraction method and PNN classifier gives the maximum accuracy amongst the all combinations with 83.75 %.

LPC features which are extracted for speech recognition contains 12 coefficients for any speech sample. These 12 coefficients for different commands like left, right, start and stop are shown in figures 4.3, 4.4, 4.5 and 4.6 respectively.

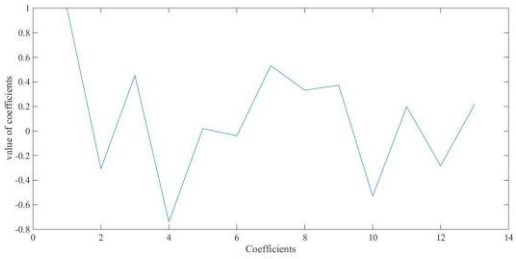



Figure 4.3: LPC for command*left*

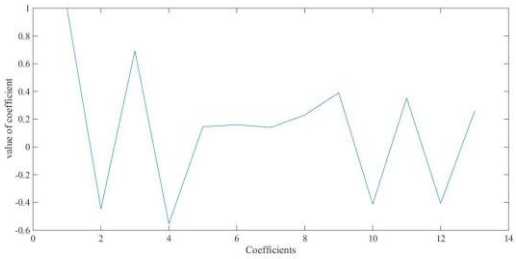



Figure 4.4: LPC for command *right*

REFERENCES

***1.*** F. Itakura, Minimum Prediction Residula Applied to Speech Recognition, IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-23(1):67-72,February 1975.

2. J. L. Ostrander, T. D. Hopmann, E. J. Delp, Speech recognition using LPC analysis, Technical Report RSD-TR-1-82, University of Michigan, 1982.

3. R. Low, R. Togneri, Speech recognition using the probabilistic neural network, Proc. 5th Int. Conf. on Spoken Language Processing, Australia, 1998.

4. Muda, L., Begam, M., &Elamvazuthi, I. (2010). Voice recognition algorithms using mel-frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. arXiv preprint arXiv:1003.4083

5. Tiwari, V. (2010).MFCC and its applications in speaker recognition. International Journal on Emerging Technologies, 1(1), 19-22.

6. Shrawankar, U., &Thakare, V. M. (2013). Techniques for feature extraction in speech recognition system: A comparative study. arXiv preprint arXiv:1305.1145**.**

7. Ittichaichareon, C., Suksri, S., &Yingthawornsuk, T. (2012). Speech recognition using MFCC. In International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July (pp. 28-29).

8.Tsilfidis, A., Mporas, I., Mourjopoulos, J., &Fakotakis, N. (2013).Automatic speech recognition performance in different room acoustic environments with and without dereverberation pre-processing. Computer Speech & Language,27(1), 380-395.

9. Shrawankar, U., &Thakare, V. M. (2013). Techniques for feature extraction in speech recognition system: A comparative study. arXiv preprint arXiv:1305.1145

10. Feature Extraction Techniques for Speech Recognition: A Review,International Journal of Scientific & Engineering Research, Volume 6, Issue 5, May-2015 ,ISSN 2229-5518

11. speech Recognition Using Deep Neural Networks:A Systematic Review, February 1, 2019,

12. Analysis of Feature Extraction Techniques for Speech Recognition System,ISSN: 2278-3075,Volume-8, Issue-7C2, May 2019