IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Vachantar –Lokbhasha: A Speech to Text Conversion for Marathi

Archana. V. Chechare[1]

P.G. Student, Computer Engineering, Sanjivani College of Engineering, Kopargaon, India [1]

Abstract: Speech processing has always been so important application area of digital signal processing. The various fields are available for researching in speech processing that are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of Vachantar Lokbhasha, A Speech to text conversion for Marathi is to recognize speaker's speech and then convert it to text. This text will be saved to a file and that file can be further used. Automatic speaker recognition is to extract and recognize the information about speaker's speech. Feature extraction is the first step for this. Many algorithms are stated by the researchers for feature extraction. In this work, the Mel Frequency Cepstral Coefficient (MFCC) feature extraction algorithm has been used for designing this system. In this Artificial Neural Networks will be used for feature classification.

Keywords: Automatic Speech Recognition (ASR), Mel-Frequency Cepstral Coefficients (MFCC), Feature Extraction, Neural Networks.

I. INTRODUCTION

Language is very efficient medium of communication. Whenever humans wants to share their thoughts, ideas anything they prefer to interact with each other by speaking to each other. So similarly for interacting with machines also human voice can be used. It is done through Automatic Speech Recognition. So by using human voice we can guide our systems for doing different things. Hence, Speech Recognition is becoming one of the most emerging fields in the computing research.

In the "Vachantar-Lokbhasha: Speech to Text Conversion for Marathi", also same technique is going to be designed. In this it is planned to develop such a system which will recognize Speech of the user and convert it to text in Marathi. Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. The acoustical parameters of spoken signal used in recognition tasks have been popularly studied and investigated, and being able to be categorized into two types of processing domain: First group is spectral based parameters and another is dynamic time series.

The most popular spectral based parameter used in recognition approach is the Mel Frequency Cepstral Coefficients called MFCC. In this paper system will focus on recognizing set of Marathi characters {अ,इ,प,औ,आ,त,क,स,ए} by using MFCC. Characters will be recognized and printed to a text file. Recognition technique makes it possible to the speakers voice to be used in verifying their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information service, voice mail, security control for the confidential information areas, and remote access to computers.

II. RELATED WORK

In the paper [2] Archit Kumar, Charu Chhabra mostly speaks about identifying and verifying speaker, where Speaker identification is mainly based on the pitch of the signal according to pitch we can distinguish that is speaker a male or a female and by using our previously stored database we can verify that is our speaker a new one or old one.

Nilu singh, R.a. Khan, Raj Shree [3] talks about techniques used for extracting features of speech and in this they have implemented feature extraction by using varying no. of filters to pass the voice and calculate efficiency. It also discusses about linear predictive coding (LPC), Linear Discriminant Bases (LDB).

Ms. Sushmita Iqbal et.al[4] have sought to build on the success in the speech recognition community by investigating how applicable it is to use the dominant features for modelling speech to model music. In this initially they discussed the process of forming MFCC feature speech, describing the reasons for and assumptions made at each step and then investigated two of the more controversial steps in the context of music modelling.

Vibha Tiwari [5] has given main focus on the differences between MFCC and Prosodic feature extraction. These two techniques are available for short term and long term feature extraction respectively. MFCC technique is mainly used for simple feature extraction in a speech whereas prosodic technique is used for extracting features for understanding and recognizing emotional status of the speaker.

In the paper T.B.Adam & M.D.Salam [6] discussed about voice sample observed with MFCC for extracting acoustic features and then used to trained HMM parameters

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

through forward backward algorithm which lies under HMM and finally the computed log likelihood from training is stored to database. It will recognize the speaker by comparing the log value from the database against the PIN code. It is implemented in Matlab 7.0 environment and showing 86.67 results as correct acceptance and correct rejections with the error rate of 13.33.Security is an essential part of human life. In this era security is a huge issue that is reliable and efficient if it is unique by any mean. Voice recognition is one of the securities.

III.PROPOSED SYSTEM

In the "Vachantar-lokbhasha: Speech to text conversion for Marathi", Speaker's voice will be processed and unique features from that will be extracted by using MFCC. Then those features will be further classified by using Neural Networks. So the whole process is divided into two sub steps those are as follows:

A. Speech signal Processing by using MFCC

Pre-processing of a signal is applying any required form of processing to the signal in time domain before the feature extraction phase .In this stage the speech signal goes through several common processes including Analog to digital (A/D) conversion, enhancement, pre-emphasis filtering and usually for SR applications silence removal or end point detection (EPD). The A/D process converts a sound wave into its digital form. There are three steps in the A/D Conversion process which are sampling, quantization and coding. The final outcome of this process is a digital version of the speech signal that can be used as well as processed by the computers. In speech recognition and processing in general, speech enhancement is conducted to suppress unwanted noise from the speech signal. For SR application removing noise increases the accuracy of the recognizer. In almost all SR application a pre-emphasis filtering step is conducted to the speech signal. The pre emphasis filter is used to emphasis the speech spectrum above 1 kHz which contains important aspects of the speech signal and equalizes the speech propagation trough air [6].

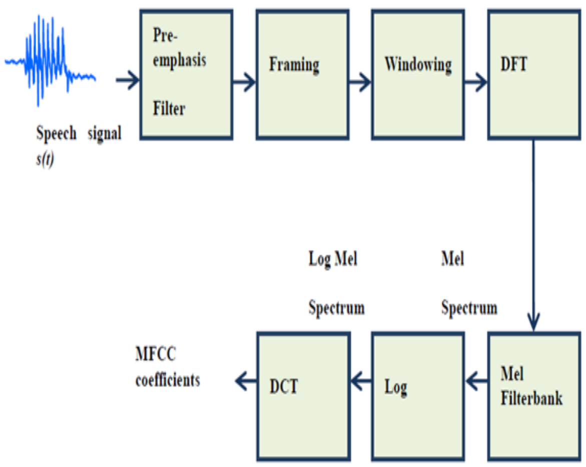

Fig 3.1 Speech Signal Pre-processing by MFCC

The Mel-frequency Cepstrum Coefficient (MFCC) technique is used for creating the fingerprint of the sound files. The MFCC is based on the known variation of the human ears Critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. It is scientifically proven that human perception of the frequency contents of sounds for speech signals does not follow a linear scale. Thus for each audio signal with an actual frequency, f, measured in Hz, a subjective pitch is measured on a scale called the Mel scale. The Mel-frequency scale is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

The basic formula developed for computing the Mels for a particular frequency is: $mel(f) = 2595*\log10(1+f/700)$. MFCC processes block diagram is shown in above Figure 1. The speech waveform is cropped to remove acoustical interference that is present in the beginning or end of the sound file. The windowing minimizes the discontinuities of the signal by tapering the beginning and end of each frame to zero. The Fast Fourier Transform converts each frame from the time domain to the frequency domain. In the Mel-frequency wrapping block, the signal is plotted against the Mel spectrum to mimic human hearing. In the last step, the Cepstrum, the Mel-spectrum scale is converted back to standard frequency scale. This Mel-spectrum provides a good representation of the spectral properties of the signal which is a key for recognizing & representing characteristics of the speaker. After the fingerprint is created, we will also going to create an acoustic vector. This vector can be stored as a reference in the database [6].

B. Classification By using Neural Networks

The second most important step in an Automatic Speech Recognition system is the classification stage. This stage includes classifying the input speech to determine whether the input speech uttered matches the desired targeted speech or not. Categories of classification schemes are statistical and artificial intelligence approaches. In this, we chose neural networks (NN) as one of the artificial intelligence approach. Neural networks (NN) are parallel distributed information processing structure with processing elements connected through unidirectional signal channels called connections. ANNs consist of simple interconnected processing elements that are called neurons that perform weighted summation of inputs [6].

IV.EXPERIMENTAL SETUP & RESULTS

A. Experimental Setup

The system is being built using Java framework (version jdk 7) on Windows platform. The Net beans (version 7.0) are alternatively used as a development tool. The system doesn't require any specific hardware to run; any standard

DOI 10.17148/IJARCCE.2016.5736

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

machine is capable of running the application. Sometimes system's detachable microphones are required for recording voice. For training this system 11 different voice samples of each Marathi character are recorded from different people in different environments.

After training phase system is tested by taking input voice signal from user in offline and runtime modes. Each mode gives different results.

B. Results

1. Offline Testing For Known Voice

TABLE I Offline Testing For Known Voice

Sr. No	Marathi Character	Recognition Result
1	अ	Correct
2	आ	Correct
3	ए	Correct
4	औ	Correct
5	इ	Correct
6	क	Correct
7	प	Correct
8	स	Correct
9	त	Correct
Result are 100% accurate(approx)		

2. Offline Testing For Unknown Voice

TABLE III Offline Testing For UnKnown Voice

Sr. No	Marathi Character	Recognition Result
1	अ	50% Correct
2	आ	Correct
3	ए	Not Correct
4	औ	Correct
5	इ	Not Correct
6	क	Not Correct
7	प	50% Correct
8	स	50% Correct
9	त	Correct
Result are 50% accurate(approx)		

3. Runtime Testing For Known Voice

TABLE IIII Runtime Testing For Known Voice

Sr. No	Marathi Character	Recognition Result
1	अ	Correct
2	आ	Correct
3	ए	Not Correct (अ)
4	औ	Correct
5	इ	Not Correct(ए)
6	क	Not Correct(अ)
7	प	Not Correct(क)
8	स	Correct
9	त	Not Correct(अ)
Result are 40% accurate(approx)		

4. Runtime Testing For Unknown Voice

TABLE IV Runtime Testing For UnKnown Voice

Sr. No	Marathi Character	Recognition Result
1	अ	Correct
2	आ	Correct
3	ए	Not Correct(त, आ)
4	औ	Not Correct(आ)
5	इ	Not Correct(आ)
6	क	Not Correct(आ)
7	प	Not Correct(आ)
8	स	Not Correct(त)
9	त	Not Correct(क)
Result are 15% accurate(approx)		

V. CONCLUSION

As in this paper it is described about designing a system for speech recognition in Marathi, different tests has been performed in Offline and Runtime modes. And hence depending on the testing environments system shows different accuracy values. The accuracy of the system in offline mode is more as compared to runtime mode.

ACKNOWLEDGMENT

The author would like to thank the researchers as well as publishers for making their resources available. Also the authorities of Sajivani College of Engineering, Kopargaon for providing the required infrastructure and support.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Finally, we would like to extend a heartfelt gratitude to friends and family members.

REFERENCES

[1] H.B.Chauhan, Prof B.A.Tanawala ,"Comparative study of MFCC and LPC algorithms for Gujrati isolated word recognition", International Journal of Innovative Research in Computer and Communication Engineering, vol-3, Feb 2015.
[2] Archit Kumar,Charu Chhabra,"Intrusion detection system using Expert System(AI) and pattern recognition (MFCC improved VQA)", International Journal of Advanced Research in Computer science Management studies,Vol-2,May2014.
[3] Nilu singh,R.a.Khan,Raj Shree, "MFCC Prosodic feature extraction techniques: A Comparitive Study", International Journal of Computer Applications(00975-8883), September 2012
[4] Shushmita Iqbql, Tahira Mehboob, Malik sikandar Hayat khiyat," Voice Recognition using HMM with MFCC for secure ATM", International Journal of Computer Science, vol- November 2011.
[5] Vibha Tiwari," MFCC and its Applications in Speaker Recognition", International Journal on Emerging Technologies 1(1):19-22,2010.
[6] T.B Adam, M.D.Salam, "Spoken English Alphabet Recognition with Mel frequency Ceptral Coefficients and Back propagation Neural Networks", International Journal of Computer Applications(0975-8807) vol-42-no12, March 2012
[7] Chadwan ittichaichareon, Siwat Suksri,Thawersak Yingthawarnsuk, " Speech Recognition using MFCC", International Journal on Computer Graphics, Simulation and Modeling(IJCGM 2012) July 2012
[8] Aldebara klatau "The MFCC", 22nd November 2005.
[9] Rohini Bhujang rao Shinde," Analyzing Childrens speech for Biometric Identification" ,International Journal of Speech and Language Processing, Vol-1, January 2011.
[10] Gopalakrishna Anumanchipalli, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Pal Singh,R.N.V. Sitaram, , "Development of Indian Language Speech Databases for large vocabulary speech Recognition System", International Institute of Information Technology, Hyderabad, Hewlet Packard Labs India, Banglore.
[11] Microsoft Cortana, https://en.m.wikipedia.org/wiki/Cortana_(software).
[12] Google voice search, https://en.m.wikipedia.org/wiki/google_voice_search.
[13] Google voice maps, http://www.idigitaltimes.com/google-maps-tips-and-trics-10-little-known-voice-commands-use-googles-navigation-493297
[14] Vachantar Lokbhasha, http://cdac.in/index.aspx?id=mc_st_vachantar
[15] Wave file format, http://soundfile.sapp.org/doc/WaveFormat/

BIOGRAPHY

Archana V. Chechare has completed her Bachelor's degree from Government College of Engineering and research, Awsari, Manchar Pune, Maharashtra, India in 2013. She is currently pursuing Master of Engineering with specialization in Computer Science and engineering from Sanjivani Rural Education Society's College of Engineering, Kopargoan, Ahmednager, India. Her current research interest includes Multilingual support in computing environment.