IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 6, Ver. II (Nov - Dec .2015), PP 58-62
www.iosrjournals.org

Speech to text conversion for visually impaired person using μ law companding

Suraj Mallik[1], Rajesh Mehra[2]
[1,2]*Department of Electronics & Communication Engineering, National Institute of Technical Teachers' Training & Research, Chandigarh-160019, India*

Abstract: *The paper represents the overall design and implementation of DSP based speech recognition and text conversion system. Speech is usually taken as a preferred mode of operation for human being, This paper represent voice oriented command for converting into text. We intended to compute the entire speech processing in real time. This involves simultaneously accepting the input from the user and using software filters to analyse the data. The comparison was then to be established by using correlation and μ law companding techniques. In this paper, voice recognition is carried out using MATLAB. The voice command is a person independent. The voice command is stored in the data base with the help of the function keys. The real time input speech received is then processed in the speech recognition system where the required feature of the speech words are extracted, filtered out and matched with the existing sample stored in the database. Then the required MATLAB processes are done to convert the received data and into text form.*

Keywords: *Asr, Dsp, Gui, Hmm, Matlab, Stt*

I. Introduction

Since the beginning of humanity and the interaction with the virtual technological world, technology has dramatically changing our life and living style. Research in Human Language Technology has made great progress in the past few years. The challenge to design much better automation processes is to accommodate the variation in between different user. Also, a unique and better user interface design can be the solution to some existing automation process design problems. Automatic speech-to-text (STT) processing systems are capable of producing English word transcripts of conversational telephone speech at 15.2% word error rate, a decrease of 53% over the past 5 years. An ideal user independent interface still does not exist at present and to build an ideal interface requires knowledge of both sociological linguistic and technological fields. According to many major companies that are involved in building speech recognition system and researches, speech will be the primary interface between humans and machines in the near future.[1] Research and development group have investigated the possibility of using speech activation in cars to enable hands free controlling. Recently, a Hidden Markov Model (HMM) based speech recognition and processing system was implemented in to enable voice activated wheelchair controlling. Speech recognition technology allows computers to translate speech in pure audio or spoken form and convert it to text format. By providing a specific grammar and limiting the vocabulary, the system needs to recognize the speech with good recognition results. The performance of the speech recognition in home environments depends on the implementation of the speech recognition system

Language is the ability to express one's thoughts by means of a set of signs, whether graphical gestural, acoustic, or even musical. It is distinctive nature of human beings, who are the only creatures to use such a structured system. Speech is one of its main components. It is by far the oldest means of communication between human being and also the most widely used. No wonder, then, that people have extensively studied it and often tried to build machines to handle it in acoustic way. Most of the Information in digital world is accessible to a few who can read or understand a particular language. Language technologies can provide solutions in the form of natural interfaces so the digital content can reach to the masses and facilitate the exchange of information across different people speaking different languages. These technologies play a crucial role in multi-lingual societies such as India which has about 1652 dialects/native languages.[2] A speech to text converter convert's normal language into text. Synthesized speech can be created by concatenating pieces of recorded speech in the form of wav. file that are stored in a database. Systems differ in the size of the stored speech units; a system that stores speech signal using microphone, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. Here question arises that whether machine or simply computer can perform same task of text to speech conversion? Answer is not that much easily as human can.[3] The machine has to follow some procedure which is divided in basic two steps: I : Speech sample recognition. Next step is STT that is speech to text conversion in this we have to convert recognized speech into text format.

II. Speech Sample Recognition

Recording of voice sample ('A,B,C,…Z') is done to improve the accuracy and preciseness of the sample to be filtered and then convert into .wav file which is MATLAB executable format to read an audio signal. Once the speech samples are stored in the system, the specified location is to be carried out using wavread command by specifying different function variables to different speech signal.[4]

1. Extracting required speech from given speech signal.

A speech word sample will be extracted from no. of samples & kept in separate word. The Microphone is used to store the input signal into the system which is then filtered and processed in MATLAB using DSP techniques.

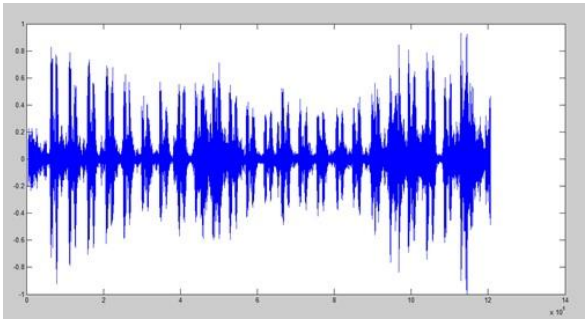

Fig.1 Input voice sample

2. Speech Analysis & Speech Detection

Speech analysis is mainly concern with analysis of extracted text from given voice samples which are in wav. file.[5] Organize and maintain them into a list of words. This list contains abbreviation, numbers, and acronyms & converts them into a full line when needed. Speech Detection is a process of identifying preciously where it is located in that given voice sample.

After recording the desired voice signal sample, the individual samples are filter which can be done using various DSP filtering techniques or carried out by extracting the required Speech signal individually through graphical method. The next thing is to correlate different speech signal with each other keeping the sample bits same. For this, the sample bits need to be extracted in such a fashion that the matrix so formed should be of the same order.[6]

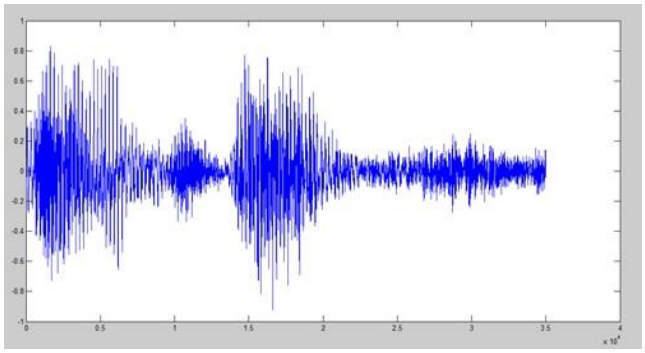

Fig.2 Individual filtered voice sample plot

3. Speech Transformation

It is normalization of speech to pronounceable from. It pronounces line by line words take pause when space is detected between words. It reads the speech according to the punctuation rules, accent marts & stop words much similar as many users. The individual samples of speech signal are then separate out from the original voice data.[7] Then by using correlation and μ law companding technique, the required signals can be synthesize in MATLAB for correlating the stored voice sample with the target one that the user will command.

R = corrcoef(X)returns a matrix R of correlation coefficients calculated from an input matrix X whose rows are observations and whose columns are variables. The matrix R = corrcoef(X) is related to the covariance matrix C = cov(X) by

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(j,j)}}. \qquad \ldots(1)$$

corrcoef(X) is the zeroth lag of the normalized covariance function, that is, the zeroth lag of xcov(x,'coeff') packed into a square array.
R = corrcoef(x,y) where x and y are column vectors is the same as corrcoef([x y]).

4. Pre-processing:

Pre-processing consists of a number of preliminary steps to make the raw data usable for recognizer. Firstly the raw input signal from the user is converted in to MATLAB executable file format as Wav. And then it is further transform into matrix form so as to compare the real time data with the stored one.[8] The noise free speech signal is passed to the segmentation step, where the individual speech signal is segmented in to characters. It is the most important aspect of pre-processing stage.

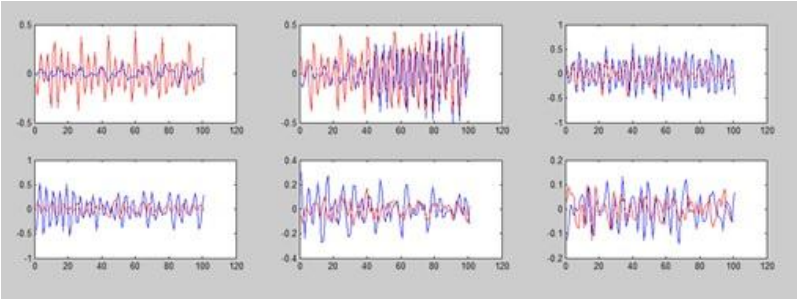

Fig.3 Comparative plot of voice samples

out = compand(in,param,v) implements a μ-law compressor for the input vector in. Mu specifies μ, and v is the input signal's maximum magnitude. out has the same dimensions and maximum magnitude as in.
out =compand(in,Mu,v,'**mu/compressor**')
The real time speech signal is then input to the system for correlate with the stored one using;
file = sprintf('%s%d.wav','rec',i);

The system is thus made to ask the user to input the user voice and then again that input voice sample is converted into the required data matrix which has to be correlate with the stored data sample.[9] A threshold value which is the average of the entire correlation output data sample is set so as to give the required output from the system as which particular speech or voice is said by the user.

III. Speech Synthesis

A Speech-To-Text (STT) synthesizer is a computer based system that should be able to read any speech aloud, when it is directly introduced in the computer by an operator.

IV. What Is Speech Synthesis?

Speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT). All speech-to-text systems rely on at least two models: an acoustic model and a language model. In addition large vocabulary systems use a pronunciation model.It is more suitable to define Speech-To-Text or speech synthesis as an automatic production of text, by given speech signal as transcripted data input to the system in real time. SR systems use "training" (also called "enrolment") where an individual speaker reads text or isolated vocabulary into the system.[10] Alpha numeric characters are the smallest distinguishing unit in a speech language. It does not carry meaning by itself. Alpha numeric characters include alphabetic letters, numerical digits, punctuation marks, and the individual symbols of any of the world's English language systems. A phoneme is "the smallest segmental unit of sound employed to form meaningful utterances" The first task faced by STT system is the conversion of input speech into text representation.
The basic types of synthesis system the following are:
• Formant
• Concatenated
• Prerecorded

1. Concatenative Synthesis:

Concatenative synthesis is based on the concatenation (or stringing together) of segments of recorded speech. Generally, concatenative synthesis produces the most natural-sounding synthesized speech. There are three main sub-types of concatenative synthesis.[11]

1.1 Unit Selection Synthesis:

Unit selection synthesis uses large databases of recorded speech. During database creation, each recorded utterance is segmented into some or all of the following: individual phones, diphones, half-phones, syllables, morphemes, words, phrases, and sentences.

1.2. Diphone Synthesis:

Diphone synthesis uses a minimal speech database containing all the diphones (sound-to-sound transitions) occurring in a language.

2. Formant Synthesis:

Formant synthesis does not use human speech samples at runtime. Instead, the synthesized speech output is created using additive synthesis and an acoustic model (physical modelling synthesis) Parameters such as fundamental frequency, voicing, and noise levels are varied over time to create a waveform of artificial speech.[12] This method is sometimes called rules-based synthesis;

3 Prerecorded Synthesis:

In prerecorded synthesis record large paragraphs of English words (commonly used English vocabulary) in a continuous rhythm with small gap between two successive words in form of a silence and save them as sound files on the database.

V. Stt Processes

1. Create a data base and call back functions.

1.1 Create a STT data base which are the user distinguish language alphanumeric characters which recognise the user speech signal when call back in real time application.

1.2 Get notified on state changes, language changes, recognition results, and errors by registered callback functions.[13]

2. Start, stop, and cancel recognition.

2.1 Start recording user voice by microphone and analyze the recorded data as text.

2.2 If you stop recording manually by API, the STT stops the recording and recognizes the sound data. Then, the recognized text is sent by the recorded callback function.

2.3 You also can set sounds which are played before the STT recording start or after the recording stops.

3. Get the recognition result.

3.1.The recognition result is sent by the required callback function.

3.2 With a specific STT engine, you can obtain the time stamp information for the recognition result.

VI. Results

A GUI is made and implemented for Speech to text conversion as shown in the following fig. The adjacent graph plot showing the given input speech signal being fed into the system in real time from the user.

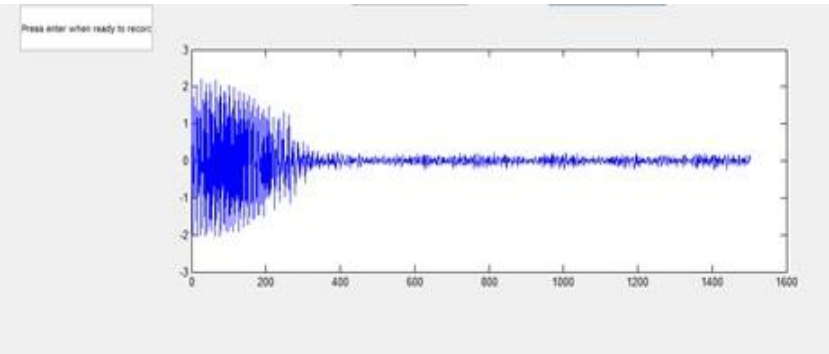

Fig. 4 Graphic User Interface

Command window showing the execution of the program in back end where the given real time signal is correlate and using μ law commanding the given input sample is compared with the stored data sample to carry out output voice as "Yes voice matched" to carry out which significant word is being said by the user.[14]

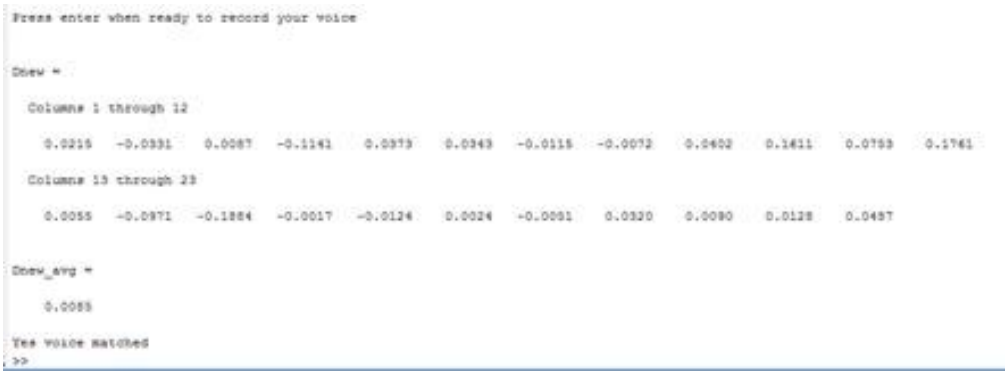

Fig. 5 MATLAB Command Window

VII. Conclusion

A system based on voice recognition was built and implemented. The system is targeted at elderly and disabled people and also to robotics applications.The proposed system therefore provides solutions for the problems faced by old or disabled persons in daily life and makes their life easier and more comfortable by proposing a cost effective and reliable solution. The system developed can be used to control AC and DC appliances through speech.The prototype developed can control electrical devices in a home or office. Confirmative voice with specific voice pitch and frequency is desired by the speech recognizer used in this system to produce better recognition results. The system controls extended and multiple appliances by using speech recognition technology. It can be applied in various applications such as voice activated wheel chairs, robotic control appliances etc.

Acknowledgements

The author would like to thank Director and Head of Electronics and Communication Engineering Department, National Institute of Technical Teachers' Training & Research, Chandigarh, India and for their constant inspirations, support and helpful suggestions throughout this research work**.**

References

[1] Poonam.S.Shetake, "Review of text to speech conversion methods" International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, pp. 28-32, Aug.-2014
[2] R. Gadalla, "Voice Recognition System for Massey University Smarthouse," M. Eng thesis, Massey University, Auckland, New Zealand, 2006.
[3] Kyung-Saeng Kim and Kwyro Lee, "Low-power and area efficient FIR filter implementation suitable for multiple taps", IEEE Transaction on Very Large Scale Integration, Vol. 11, no. 1, pp.150-153, 2003.
[4] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, New Jersey, US: Prentice Hall Inc, 1978.
[5] Devendra Kumar Somwanshi, Image Acquisition, Recognition & Speech conversion, Thapar University, Patiala,M.E, july-2009
[6] Xiaohua Zeng ,Fapojuwo, A. ; Davies, R.J., Design and performance evaluation of voice activated wireless home devices, Research in Motion, Ottawa, Ont.
[7] www.mathworks.in/help/comm/ref/compand.html
[8] www.mathworks.in/help/matlab/ref/corrcoef.html
[9] V. Shanmughaneethi; Ra. Yagna Praveen; S. Swamynathan, Detection of command injection vulnerabilities in web services through aspect-oriented programming, International Journal of Computer Applications in Technology (IJCAT), Vol. 44, No. 4, 2012
[10] R. Puviarasi, Mritha Ramalingam, Elanchezhian Chinnavan, Low Cost Self-assistive Voice Controlled Technology for Disabled People, International Journal of Modern Engineering Research (IJMER,,) ISSN: 2249-6645, Vol. 3, Issue. 4, pp-2133-2138, Jul.-Aug. 2013
[11] Y Bala Krishna, S. Nagendram, Zigbee Based Voice Control System For smart Home, Int.J.Computer Techology & Applications, Vol 3 (1), 163-168 IJCTA, 163 ISSN:2229-6093, JAN-FEB 2012
[12] Monica Singhal, Rajesh Mehra, Analyzing Aliasing effect in Down Sampler with increase in factor M, International Journal of Scientific Research Engineering & Technology (IJSRET) ISSN: 2278–0882
[13] Rajesh Mehra, Shaily Verma, FPGA Based Design of Direct Form FIR Polyphase Interpolator for Wireless Communication, International Journal of Electrical Electronics &Telecommunication Engineering, ISSN:2051-3240, Vol.44,Issue.1
[14] M. AL-Rousan, K. Assaleh, "A wavelet and neural network based voice system for a smart wheel chair control" Journal of the Franklin Institute, Volume 348, Issue 1, Pages 90–100, February 2011