© 2017 IJSRSET | Volume 3 | Issue 3 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

An Alternative Voice Communication Aid based on ASR

Hritika Rathore, Jyotsna Sagar
ECE Department, SSSUTMS, Sehore, India

ABSTRACT

This paper include the implementation aspects of alternative communication aid for the person suffering from speech impairment problem. The Alternative Communication Aid (AVCA) is an automatic voice recognition system that uses the concept of automatic speech recognition system (ASR) and announce the constructed sentence. For feature extraction from sound waves we have used mel frequency cepstral coefficient (MFCC) and DTW algorithm to find the distance between two vectors. This system also incorporate the feature of record voice samples in runtime.

Keywords: MFCC, ASR, DTW, Voice Communication Aid

I. INTRODUCTION

Automatic speech recognition (ASR) is a biometric identification methodology to recognize any person. The ASR has various advantages over other biometric identification and so have involvement in various application of engineering. Alternative Voice Communication Aid (AVCA) is one of the application of ASR that is helpful for the patients of speech impairment. The AVCA is a system that take impaired voice as input and generate a voice output for that specific speech input. AVCA uses the concept of voice recognition to detect the word spoken in impaired voice and then show a predicted response to the user for constructing the sentence to be announce by the AVCA system.

II. METHODS AND MATERIAL

A. Mel Frequency Cepstral Coefficient (MFCC)

The phonemes is one of the unit of sound by virtue of which two words can be distinguished by each other. This can only be accomplished by determining the vocal track shape of any person. The main part that has to be understand carefully is that the speech that generated by humans are filtered by their vocal track that includes teeth, tongue, etc. Before understanding MFCC clearly we first discuss about MEL SCALE and CEPSTRAL of Signal.

- **MEL SCALE**, The word MEL is taken form word melody that indicates the scale is pitch comparison based. The MEL scale has been introduced by Newman Volksman and Stevens in 1937.

It is a perceptual scale which is judged by listener to be in equal distance from each other. There is a reference point assigned in perceptual pitch 1000 mel to 1000 Hz tone in frequency measurement scale with 40dB and above the listener's threshold. Mathematically there is no any standardized formula for calculating mel of any frequency component. A commonly used formula for this is given below.

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

Here m = value in mel
F = value in Hz.

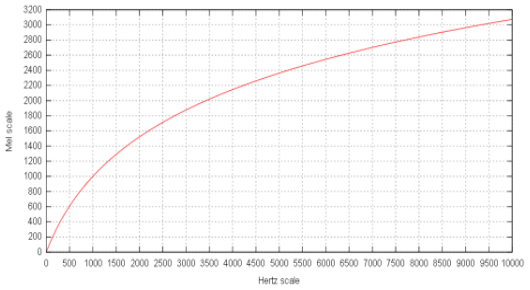


Figure 1. Graph between signal in mel value and freq value

- **CEPSTRUM COEFFICIENT,** It is the result that comes after the Inverse Fourier transform of Log of estimated spectrum of any signal. Cepstrum

coefficients are of four different type first is Power Cepstum useful in human voice recognition system, second is Complex Cepstrum uses complex logarithmic functions that holds some magnitude and phase information, Third is Real Cepstrum uses a logarithmic function that only holds magnitude information and last is Phase Cepstrum that holds phase information of initial spectrum.

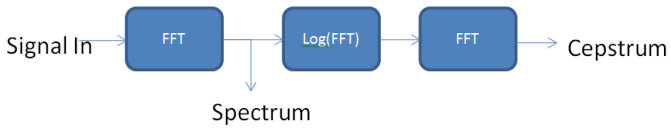


Figure 2. Block diagram of Cepstrum Coefficient Calculation

The block diagram of MFCC is shown in figure 3. MFCC is a parameterization technique that incorporate various calculation steps to detect the shape of vocal track.

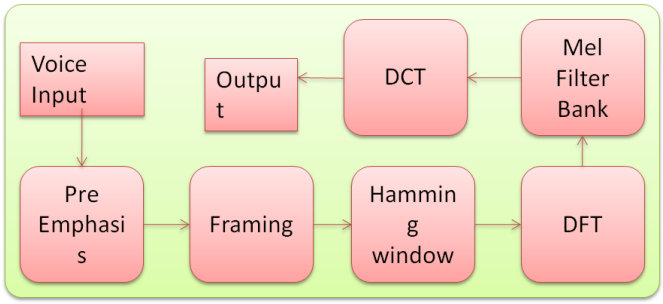


Figure 3. MFCC Block Diagram

Voice Input: The recoded signal are sampled in this block and this sampled voice signal is then applied to Pre Emphasis block.

Pre-emphasis: This block of calculation spectrally flatten the signal for the sack of this a pre-emphasis filter is used.

Framing: This block of calculation is responsible to segment analog data into small size frames 20mSec.

Hamming Window: The small sized frames are truncated in this block with the help of Hamming window function.

DFT: This block converts the time domain signal into frequency domain for further calculations.

Mel Filter Bank: This block has a filter bank been created by calculating the number of peaks in the signal that are uniformly placed in mel scale.

DTC: This block is responsible to convert frequency domain to time domain. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

B. Dynamic Time Wrapping (DTW)

In order to recognize any word that has been spoken by user is to record that word and then compare it with available matching words stored in database. To find the best suited word DTW algorithm can be use. Let us assume we want a word or command to be recognize. A word W of Tx vector long with sequence of vectors as $\tilde{X}=\{\vec{x}_0,\vec{x}_1,\ldots,\vec{x}_{(X-1)}\}$ from acoustic pre-processing stag. Now the task is to calculate distance between $\tilde{X}$ and known vector sequences $\tilde{W}_k=\{\vec{w}_{(k0,)}\ \vec{w}_{(k1,)}\ \ldots,\vec{w}_{(kT_{(W_k)})}\}$. Let us consider that V different words w_0,w_1,...w_(V-1) are stored in our vocabulary (a set of Ω-classes). We can use a word w_v (here: class w_v□Ω) that represent a set of prototypes $\tilde{W}_{(\llbracket k,\omega \rrbracket_v),k=0,1,\ldots,(K_{(\omega_v)}-1)}$ in order to classify using Nearest Neighbour Classification task. This will represent all possible variations due to different pronunciation/speakers.

III. IMPLEMENTATION, WORKING AND RESULT

The schematic representation of the system is shown in figure. There are main four parts of the system and is build around these four major blocks.

A. Implementation

Speech Recording and Recognition: This block of system is responsible to record the voice signals and then compare it with existing voice samples in the database. For recognition of voice we used MFCC and DWT algorithm together to find the best result from pre-recorded samples of voice.

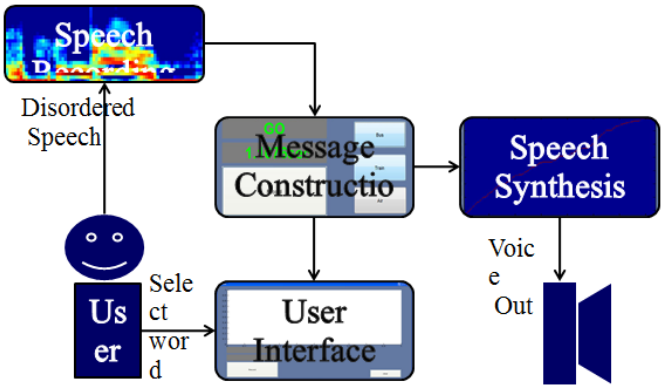


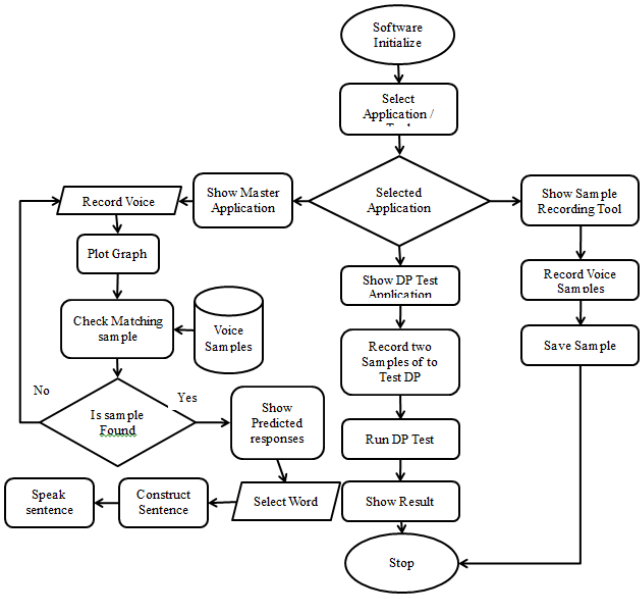
Figure 4. Schematic View of AVCA System

User Interface: As the user need to record voice samples and then construct message to be spoken by AVCA system. A user interface is designed for the this.

Message Construction: This is the main part of our system as user have to select one of the predicted response given by system. As user select one option from given words system hold the response and then show next level predicted response.

Speech Synthesis: In this part the constructed message is then announced by the system.

B. Working and Result



The flow chart of the application is shown in figure 5. The complete application is divided into three parts first is master application used for recognition and announcement system, second is DP test application used to check the DT algorithm response for two samples and last is Recorder tool this application can be used to store new samples into the database or to update the existing samples in our database. Figure 6 shows the main page of application all tools are connected linked on this page with the means of buttons of main page. User can select any of the application or tool to perform required task .

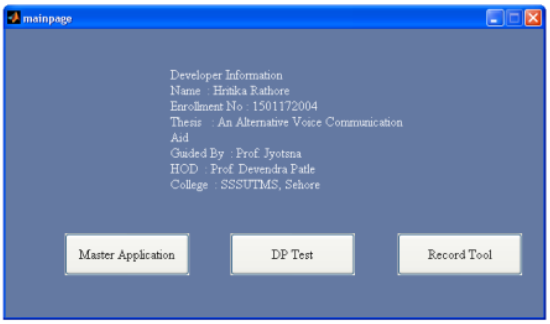


Figure 6. Main Page of Application

As user click on Master application master application from will display which is shown in figure 7. This is main application which performs the specific recognition task and show the predicted response of the speech signal that has been recorded by user by clicking the button Record.

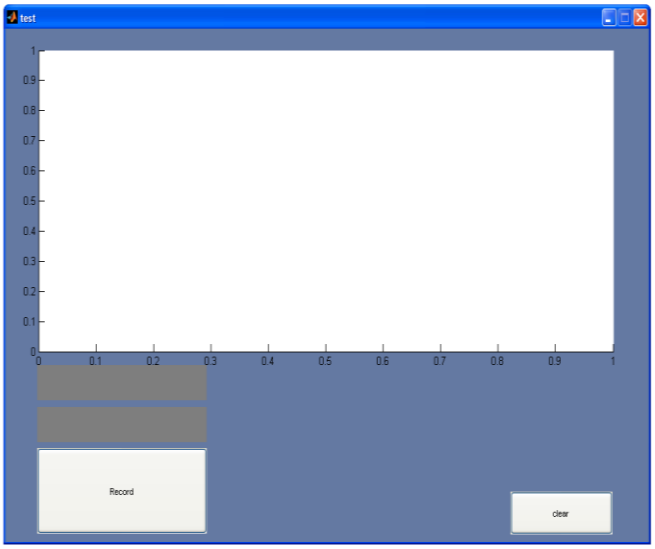


Figure 7. Master Application Window

The recorded voice is used to recognize the word which is done by DT algorithm whose output is shown in the figure 8.

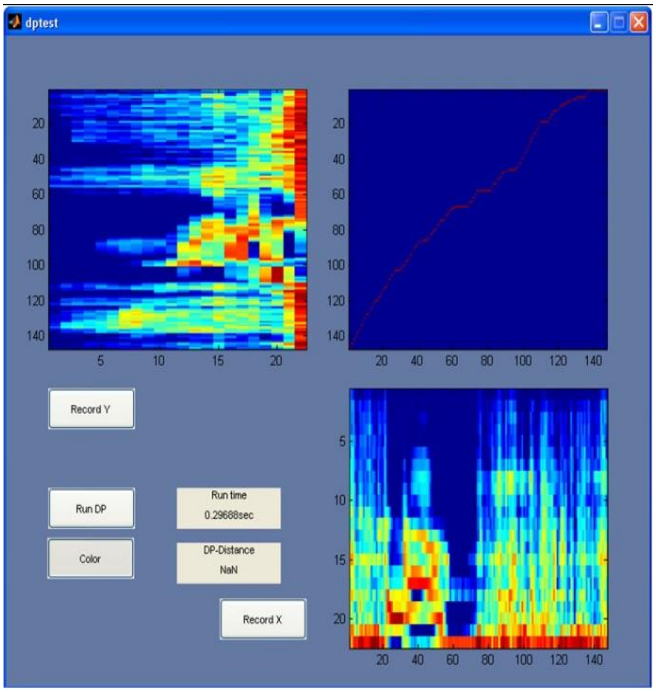


Figure 8. DT Algorithm Result

IV.CONCLUSION

The problem of speech impairment is for dysarthria patients is a big concern of our project which has been successfully accomplished. The AVCA has successfully

developed which provide a predicted result for spoken word and so one can generate a message with the help of this and then the system will announce the generated message. This system is not capable of speech recognition but it also have a feature to record new voice samples for the user.

V. REFERENCES

[1] Zhanyu Ma, Hong Yu, Zheng-Hua Tan And Jun Guo, "Text-Independent Speaker Identification Using the Histogram Transform Model", IEEE Access, VOLUME 4, 2016, pp(9733-9739)

[2] SadaokiFurui, 50 years of Progress in speech and Speaker Recognition Research , ECTI Transactions on Computer and Information Technology,Vol.1. No.2 November 2005.

[3] K.H.Davis, R.Biddulph, and S.Balashek, Automatic recognition of spoken Digits, J.Acoust.Soc.Am., 24(6):637-642,1952.

[4] H.F.Olson and H.Belar, Phonetic Typewriter , J.Acoust.Soc.Am.,28(6):1072-1081,1956.

[5] D.B.Fry, Theoritical Aspects of Mechanical speech Recognition , and P.Denes, The design and Operation of the Mechanical Speech Recognizer at Universtiy College London, J.British Inst. Radio Engr., 19:4,211-299,1959.

[6] J.W.Forgie and C.D.Forgie, Results obtained from a vowel recognition computer program , J.A.S.A., 31(11),pp.1480-1489.1959.

[7] J.Suzuki and K.Nakata, Recognition of Japanese Vowels Preliminary to the Recognition of Speech , J.Radio Res.Lab37(8):193-212,1961.

[8] T.Sakai and S.Doshita, The phonetic typewriter, Information processing 1962 , Proc.IFIP Congress, 1962.

[9] K.Nagata, Y.Kato, and S.Chiba, Spoken Digit Recognizer for Japanese Language , NEC Res.Develop., No.6,1963.

[10] T.B.Martin, A.L.Nelson, and H.J.Zadell, Speech Recognition b Feature Abstraction Techniques , Tech.Report AL-TDR-64-176,Air Force Avionics Lab,1964.

[11] T.K.Vintsyuk, Speech Discrimination by Dynamic Programming , Kibernetika, 4(2):81-88,Jan.-Feb.1968.

[12] C.C.Tappert,N.R.Dixon, A.S.Rabinowitz, andW.D.Chapman, Automatic Recognition of Continuous Speech Utilizing Dynamic Segmentation, DualClassification, Sequential Decoding and Error Recover , Rome Air Dev.Cen, Rome, NY,Tech.Report TR-71-146,1971.

[13] F.Jelinek, L.R.Bahl, and R.L.Mercer, Design of a Lingusistic Statistical Decoder for the Recognition ofContinuous Speech , IEEE Trans.InformationTheory,IT- 21:250-256,1975.

[14] F.Jelinek, The Development of an ExperimentalDiscrete Dictation Recognizer , Proc.IEEE,73(11):1616- 624,1985.

[15] GEMMA HORNERO, DAVID CONDE, MARCOS QUÍLEZ, SERGIO DOMINGO, MARÍA PEÑA RODRÍGUEZ, BORJA ROMERO, AND OSCAR CASAS, "A Wireless Augmentative and Alternative Communication System for People With Speech Disabilities", IEEE Access, VOLUME 3, 2015, PP 1288-1297

[16] Deepak Joshi, Shiv Dutt Joshi, “Improved Language Identification Using Sampling Rate Compensation & Gender Based Language Models For Indian Languages”, 978-1-4673-6190-3/13/$31.00 ©2013 IEEE

[17] Mark S. Hawley, Stuart P. Cunningham, Phil D. Green, Pam Enderby, Rebecca Palmer, Siddharth Sehgal, and Peter O’Neill, “A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment”, IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 21, NO. 1, JANUARY 2013, Pg 23-31