International Journal of Scientific Research in Computer Science and Engineering
Vol.4, Issue.3, pp.7-10, June (2016)

Research Paper

E-ISSN:2320-7639

A Predictive Voice-Input and Output Alternative Communication System using MFCC & DTW

S. Pal[1], V.P. Singh[2]

[1]Department of ECE, SSSIST, Sehore, India
[2]Department of ECE, SSSIST, Sehore, India

Available online at www.isroset.org

Received: Apr/16/2016, Revised: May/02/2016. Accepted: May/25/2016, Published: Jun/30/2016

***Abstract*—** Modern world have lots of applications that are related to speech recognition, and so a lots of researches have been carried out in this field which empower the world more and more automated. In this paper another kind of application has been designed for peoples suffering difficulty in speaking the words properly. The system records the impaired speech and takes it as input for further processing. As per input of the speech system display predictive suggestions and then finally complete sentence will be spoken by the system. Thus it is a voice input voice output predictive system.

***Keywords*-** ASR; Mel Frequency; Cepstrum Coefficient; Feature & Vector Space; Distance Measurement

I. INTRODUCTION

A person that is suffering problem of dysarthria (in which a person is suffering speech impairment) needs some sort of Augmentative and Alternative Communication (AAC) device thus they can easily communicate with other people around them. The Voice input and output alternative communication system (VIOACS) is described. The VIOACS system is capable of recognizing the impaired or disorder speech of the user and allow user to select one of predictive responses suggested by system and then system build message and announces the appropriate message. The system development is carried out employing the non-user-centered development method (i.e. a speaker independent). The experiment showed that this method is doing well in generating good recognition performance (mean accuracy 98%) in disordered speech.

II. MATERIAL AND METHODS

A. *Mel Scale*

The name for mel scale is come from melody that indicates the scale is pitch comparisons based. This scale is being named by Newman, Volksman and Stevens in 1937.

This is a perceptual scale of pitch which is judged by listener to be in equal distance from one another. There is a reference point in assigned as perceptual pitch 1000 mels (in mel scale) to a 1000 Hz tone (in frequency measurement scale) with 40dB above the user's (listener's) threshold. There is no standard formula for converting frequency into mels yet a popular formula is shown below. The figure 1 is the graph plot between signal in mel value and freq value.

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

Here m = value in mel

F = value in Hz.

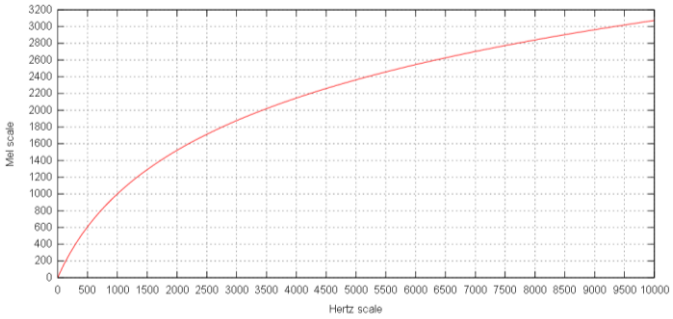

Figure 1. Graph plot between signal in mel value and freq value

B. *Cepstrum Coefficients*

The name cepstrum is came from reversing the four laters of "spectrum", CEMSTRUM. Cepstrum is the result that comes after taking Inverse Fourier Transform of log of estimated spectrum of any signal. There are four different types of cepstrum. First is Power cepstrum, mainly used in applications like human voice recognition system and etc. second is real cepstrum, Second is complex cepstrum, this type of cepstrum used a complex logarithmic function which holds the magnitude and phase information of initial spectrum. Third is real cepstrum use a logarithmic function which can only hold the information about the magnitude of initial spectrum. Fourth is phase cepstrum holds the phase details of initial spectrum.

© 2016, IJSRCSE All Rights Reserved

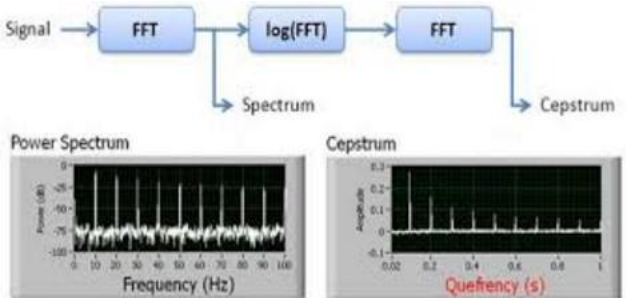

Figure 2. Power Spectrum and Cepstrum of Signal

C. *Mel Frequency Cepstral Coefficient (MFCC)*

The main thing to understand in speech is that, a sound generated by human is filtered by the shape of its vocal tract that includes teeth, tongue etc. This shape of vocal track determines what sound come out. If the shape of vocal track can be determine accurately, then the phonemes (is one of the units of sound that distinguish one word from another in a particular language) can be found.

The vocal tract shape manifests in the short time power spectrum of speech envelop, and MFCC is capable of detecting that envelope. MFCC is the one of the widely used parameterization technique. The block diagram shown in figure 3 defines the steps that has to be taken to calculate the MFCC,

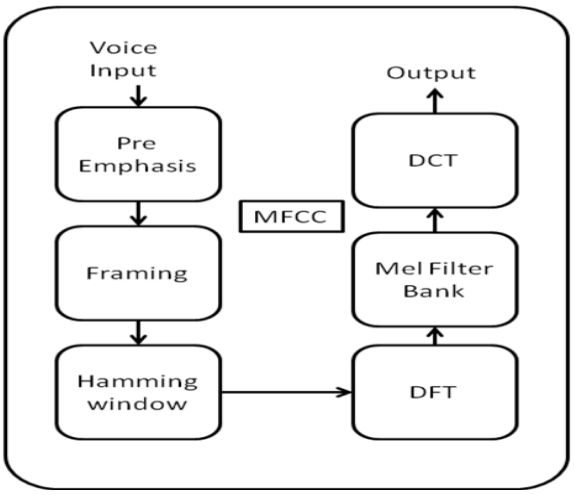

Figure 3. MFCC Block Diagram

Description of each block:

- Pre-emphasis, for making spectrally flatten signal a pre-emphasis filter is used in this step.
- Framing, in this stag analog data is segmented in small size frames of size 20mSec.
- Hamming Window, the frames are then truncated with hamming window function.
- DFT, Convert all each frame form time domain into frequency domain.
- Mel Filter Bank, in this section a bank of filters has been created by calculating a number of peaks, that are placed uniformly spaced in Mel-scale.
- DCT, Converting Freq domain to time domain

D. *Dynamic Time Wrapping*

In the field of voice recognition the ideal and simplest way is record a signal and the compare it against a number of stored word in templates and determine best matching word. Dynamic Time Wrapping (DTW), an algorithm after applying which a system can calculate similarities.

III. SIMULATION AND RESULT

A. *Simulation*

The Voice input and output alternative communication system (VIOACS) flow chart is shown in figure 4 below. As the system start it waits for record button to be pressed by user, and then record voice sample. After recording the voice sample plot calculate the MFCC of the signal, plot this signal in a graph and then check the sample from the database. The voice sample is checked by DTW algorithm.

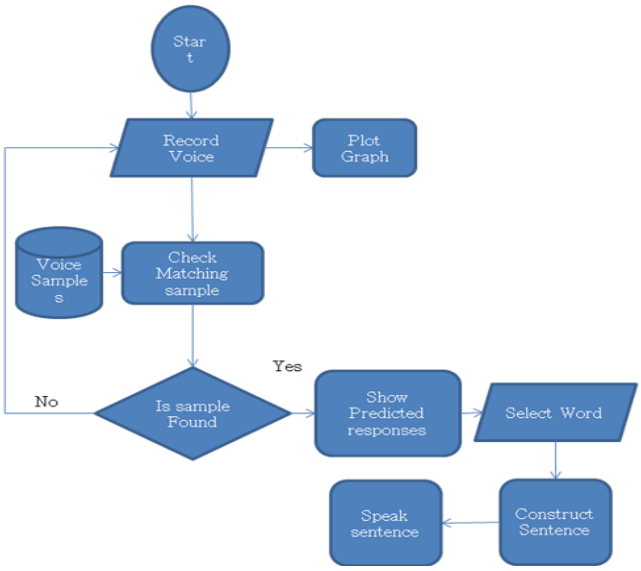

Figure 4. Flow Chart of VIOACS System

If any recorded sample is matching from the samples store in database, then show predicted responses. User now has to select the word he wants to speak, for speaking the sentence user has to construct a sentence then finally speak the sentence.

- Speech Recognition, The most common method used for recognition of voice is statically model based (usually HMM). This kind of ASR uses large number of voice samples. Yet it's a more common method for ASR but this method does not suit for recognizing disordered speech, because of limited speech samples.

© 2016, IJSRCSE All Rights Reserved

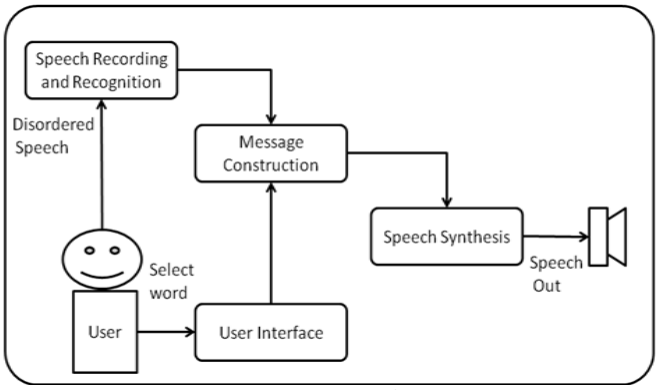

Figure 5. Schematic Diagram of VIOCS System

- Message Construction and GUI, uses GUI to select word before construction the message. As the user select message word appropriate message will construct and given to next process.
- Speech Synthesis is the process of generating human voice by machine.

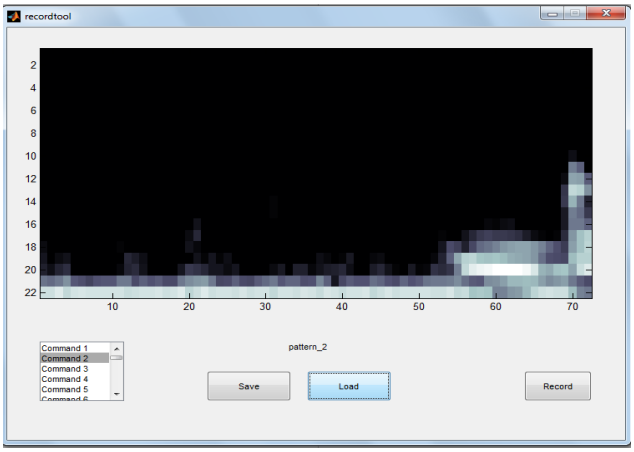

Figure 6. Voice Sample Recorder Window

B. *Result*

The final prototype of VIOACS system has been evaluated in user-trial design for assess the controllers of the VIOACS system. There is an additional window has been developed to record and store the data in of speech in database, figure 6 shows controls of recorder window. We can see the plots of signal while recoding and can store this for selected command.

The main application window is shown in figure 7, as user click on record button system goes in recoding and further calculation unit and show the matched word and time of speech recorded appropriate block, as the samples are calculated the recorded signal is plotted in graph box in the bottom. If any word match is found system shows the predicted responses related to spoken word and then user can construct sentence and synthesizer speak the sentence. Figure 8 shows the results of word "GO" matched user can select any of the button for constructing the sentences.

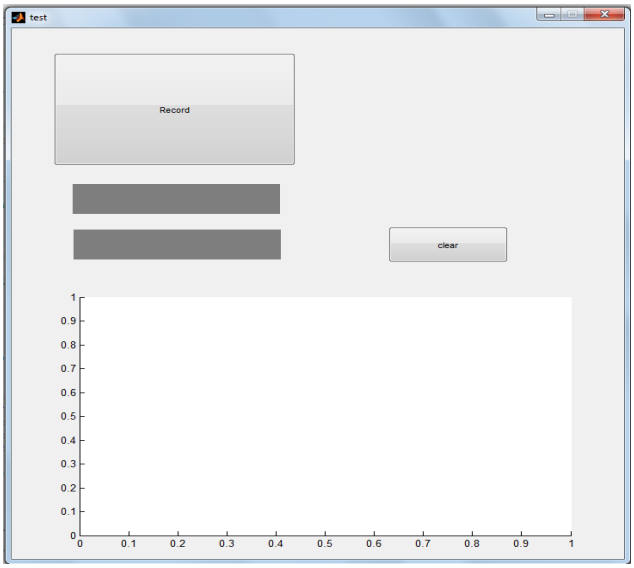

Figure 7. VIOACS System Main window

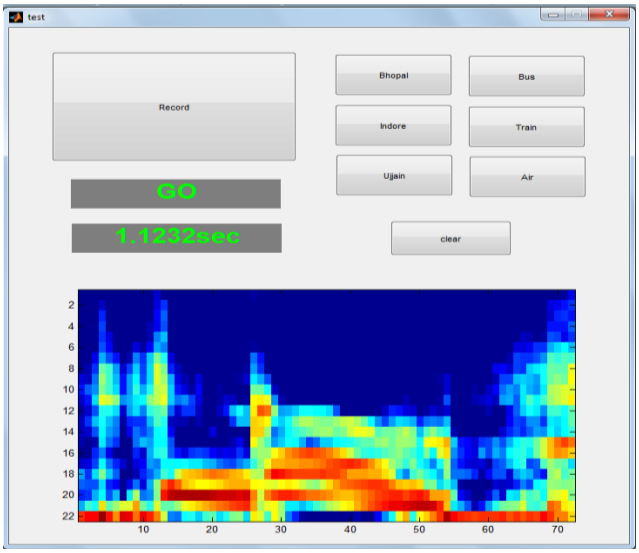

Figure 8. VIOACS Output Screen showing output for word "GO"

IV. CONCLUSION

The speech/voice communication is one of the most essential communication medium. VIOACS system has been successfully designed for a person suffering dysarthria. We have been inspired by dysarthria patents to work in this field. Researchers are attracted in speech/voice recognition to provide artificial intelligence to computer system and make it more familiar with user. Researchers are interested to make machine capable of work like human and so the speech recognition and synthesis is the dominating side of artificial intelligence system. The VIOACS system has been successfully developed as user independent system. The main objective behind developing this system is to give an interactive solution for dysarthria patients.

© 2016, IJSRCSE All Rights Reserved

REFERENCES

[1] S. Furui, "*50 years of Progress in speech and Speaker Recognition Research*", ECTI Transactions on Computer and Information Technology,Vol.1. No.2 pp.64-74, 2005.

[2] K.H. Davis, R. Biddulph, S. Balashek, "*Automatic recognition of spoken Digits"*, Journal of Acoust. Soc.Am., Vol.24, Issue.6, pp.637-642,1952.

[3] H.F. Olson, H. Belar, *"Phonetic Typewriter"*, Journal of Acoust Soc Am., Vol.28, Issue.6, pp.1072-1081,1956.

[4] D.B. Fry, *"Theoritical Aspects of Mechanical speech Recognition , and P.Denes"*, Journal of the British Institution of Radio Engineers, Vol.19, Issue.4, pp.211-8, 1959.

[5] J.W. Forgie, C.D. Forgie, *"Results obtained from a vowel recognition computer program",* The Journal of the Acoustical Society of America, Vol.31, Issue.11, pp.1480-1489.1959.

[6] J. Suzuki, K. Nakata, "*Recognition of Japanese Vowels Preliminary to the Recognition of Speech*", J ournal of Radio Research Lab, Vol.37, Issue.8, pp.193-212,1961.

[7] T. Sakai, S. Doshita, "*The phonetic typewriter*", IFIP Congress, Vo.445, pp.449-458, 1962.

[8] K. Nagata, Y. Kato, S. Chiba, "Spoken Digit Recognizer for Japanese Language", NEC Research Development, Vol.8, No.6, pp.347-352, 1963.

[9] T.B. Martin, A.L. Nelson, H.J. Zadell, "*Speech Recognition b Feature Abstraction Techniques*, Air Force Avionics Lab, US, pp.1-136, 1964.

[10] T.K. Vintsyuk, "*Speech Discrimination by Dynamic Programming*", Kibernetika, Vol.4, Issue.2, pp.81-88, 1968.

[11] H. Sakoe, S. Chiba, "*Dynamic programming algorithmoptimization for spoken word recognition"*, IEEE Transaction Acoustics, Speech, Signal Procssing, Vol.26, Issue.1, pp.43-49,1978.

[12] D.R. Reddy, "*An Approach to Computer Speech Recognition by Direct Analysis of the Speech Wave*", Technical Report No.C549 (Stanford University), USA, pp.1-173, 1966.

[13] V.M. Velichko, N.G. Zagoruyko , "*Automatic Recognition of 200 words*", Internatioanl Journal of Man-Machine Studies, Vo. 2, pp. 223-234, 1970.

[14] H. Sakoe, S. Chiba, "*Dynamic Programming AlgorithmOptimization for Spoken Word Recognition"*, IEEE Transaction Acoustics, Speech, Signal Procssing, Vol.26, Issue.1, pp.43-49,1978.

[15] F. Itakura, "*Minimum Prediction Residula Applied toSpeech Recognition"*, IEEE Transaction Acoustics, Speech, Signal Procssing, Vol.23, Issue.1, pp.67-72, 1975.

[16] C.C. Tappert, N.R. Dixon, A.S. Rabinowitz, W.D. Chapman, "*Automatic Recognition of Continuous Speech Utilizing Dynamic Segmentation, DualClassification, Sequential Decoding and Error Recover*", Rome Air Dev Cen, Rome, NY, pp.71-146,1971.

[17] F. Jelinek, L.R. Bahl, R.L. Mercer, "*Design of a Lingusistic Statistical Decoder for the Recognition ofContinuous Speech*", IEEE Transaction Information Theory, Vol.21, pp.250-256, 1975.

[18] F. Jelinek, "*The Development of an ExperimentalDiscrete Dictation Recognizer"*, Proceeding of IEEE, Vol.73, Issue.11, pp.1616- 624, 1985.

[19] L.R. Rabiner, S.E. Levinson, A.E. Rosenberg, J.G. Wilpon, "*Speaker Independent Recognition ofIsolated Words Using Clustering Techniques*, IEEE Transaction Acoustics, Speech, Signal Procssing, Vol.27, pp.336-349, 1979.

© 2016, IJSRCSE All Rights Reserved