IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 8, Issue 5, Ver. I (Sep. - Oct. 2018), PP 01-08
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Communication Aid for People With Severe Speech Disability

Prof. S.S. Bhabad, Miss. Anjali Patil
Assistant Professor Dept. Of E&TC , Pune University K.K.Wagh College Of Engineering, Nashik, Mahrashtra
ME Student Dept. of E&TC, Pune University .K.Wagh College Of Engineering, Nashik, Maharashtra.
Corresponding author: Prof. S.S. Bhabad

Abstract – *Verbal communication is a vital element in quality of life, however upwards of 1.4 percent of human beings cannot utilize natural speech reliably to communicate their views and feelings with others which leads to speech drawbacks. Speech disabilities or speech impairments are the parts of communication disorders in which the normal speech get disrupted like stuttering, lips, etc. The word disability can avert those people who are suffering from severe speech disabilities from communicating in a way of doing thing that allows them to use for one ends their potential in education & recreation. In this study a new form of speech recognition system is developed which recognizes the disordered speech of the people who are suffering from severe speech disabilities. In this work, the MFCC is used for feature extraction to extract the features of voice samples & KNN classifier is used for classification purpose. On exceptionally scattered discourse, notwithstanding when acknowledgment perplexity is expanded. Certain advantages are intelligible speech output can be produced and this method is suitable for impaired speech.*

articulation, voice and fluency. Articulation disorders means un-production of speech sounds or an absence of speech sounds. Voice disorders means in which quality of voice is affected. It is related to pitch including loudness. Fluency disorders are also known as disfluency. Fluency means repetition in speech sounds which interrupt the person's flow of the speech. There are some causes of speech impairment such as brain damage, malfunction of respiratory, stuttering, etc. Speech impairment disorder is also associated with most of other physical disabilities such as motor neuron disease, cerebral palsy, brain injury, etc [2].

Nowadays, users need a device or a communication system which is easy to handle [3]. ASR systems work well for the people who are suffering from severe speech disabilities such as dysarthria which is the most common speech disorder

[4] and [5], these studies shows that there is inverse connection between level of weakness and exactness of speech recognition. This system describes the development of speech recognition system which recognizes the disordered speech.

Index Terms- *Speech recognition, MFCC, KNN classifier, Raspberry pi 3*

Date of Submission: 04-10-2018 Date of acceptance: 16-10-2018

I. Introduction

Speech is the natural and highly interactive way of communication among humans. The general population with serious discourse issue knows about what they might want to state however unfit to express their opinions. This may lead to the development of depression. The speech of people who are the patient suffering from moderate to severe vocal disorder called as dysarthria this vocal disorder has affected about 170 per 100000 of population of the country [1]. Speech subsystems such as articulation, resonance, respiration get affected in intelligibility, audibility in vocal communication. Because of articulatory shortages, there is no consistency in the pronunciation. The pronunciation fluctuates because talking rate is moderate. The word disability can avert those people who are suffering from severe speech disabilities from communicating in a way of doing thing that allows them to use for one ends their potential in education & recreation. Speech disability is an abnormal speech i.e. Unintelligible, unpleasant or interferes with communication. There are three causes of speech impairment as

II. Literature Survey

A Hidden Markov Model was constructed [6] by using three feature vectors such as FFT, LPC and MFCC. In this speaker dependent system was investigated. A 10 state ergodic model constructed using 15ms frames. A small set of 25 isolated words was used which is the limitation of that system. MFCC which is the most knowing method of feature extraction provided WRA of 92%. Also the LPC and FFT methods provided 79% and 85% of word recognition accuracy. Another studies developed a limited database application [7] in which HTK toolkit was used which is based on HMM model with 11 states. In this MFCC (12 mel. freq. coeff.)

derived from 26 channel filter bank with 25ms analysis window with 10ms frame rate. Results of studies such as the recognition rate of preparation phase increased from 88.5% to 95.4%. Speech controlled ECS less accurate but faster to use than other scanning systems.

In one of the study an application of 10 state ergodic hybrid structure [8] constructed as handheld system for dysarthric disordered speech recognition system. In this an

HMM/ANN hybrid structure used (10 state) with MFCC (speech signal which is sampled at 11 KHz) coefficients which are extracted from 15ms frames with the window size 30ms.System performance affected by other environmental noise. Digit data provided high recognition accuracy than word. This system provides high accuracy for dysarthria data which is most common speech disorder. A new form of VIVOCA device was constructed [9] for the people who are suffering from highly speech disorders. VIVOCA catches disordered speech of the speaker & construct the messages from the recognized speech which are converted to synthetic speech. An HMM hybrid structure used with 11 states and Baun Welch algorithm was used with MFCC method for feature extraction (26 channel filter bank with 25ms analysis window), 10ms frame rate were used. It limits the performance in real usage situation. This communication device provides good recognition performance about 96% of accuracy.

A brief survey on speech recognition was presented to provide technical perspective and progress in the field of communication [10]. The study shows the comparison of known techniques which are used in the different stages of speech communication system. A speech recognition system was investigated using LPC feature extraction technique[11]. Results show that the WRA increases as increasing the value of coefficients. In one of the study shows the detailed description of feature extraction technique using MFCC which is the most commonly used in the ASR systems [12]. Results shows the effect of normalization, down sampling and changes in the parameters like linear spacing, window size. Classification was done by using minimum distance classifier. More accuracy achieved for down sampling than normalization.

One of the representations learning system for the people who are suffering from vocal weaknesses due to dysarthria converts disordered verbalization to synthesized verbalization or text was investigated [13] which recognize the sequential patterns of the varying length utterance. In this study ESHMM i.e. Example Specific Hidden Markov Model which is based on HMM constructed per class by using 5 to 10 utterance. Word recognition accuracy of Example specific hidden markov model gives better results than conventional HMM & DNN model. SR which is based on HMM which gives most flexible approach. Fundamental progress of STT conversion based on Raspberry Pi was constructed [14] in which online STT engine used. Online STT engine were used by using linux commands so that the speech signal easily processed and recognizes the text. Previous studies shows that the repeated practice can get better effect [15] in the stabilizing the target utterance.

By referring all the previous studies, we have developed a speech recognition system for the people who are suffering from severe speech disorders. This system can recognizes the disordered speech with maximum accuracy than the previous investigation.

III. System Description

Speech signal is an efficient method of communication. And because of this researcher making observations have thought of using speech for effect on one another between machine and man as it provides higher rate of efficiency. Speech is a continuous time varying signal. In this study we have developed a speech recognition system which recognizes disordered speech of the people who are suffering from severe speech disabilities. Speech recognition algorithms divided into two types such as speaker dependent and speaker independent. The goal of speaker dependent framework is to

constructing a system to acknowledge exceptional voice samples of people. Speaker independent system includes recognizing the word articulated by the speaker. Furthermore, it can be again divided into unique word detection and continuous speech acknowledgment [16].

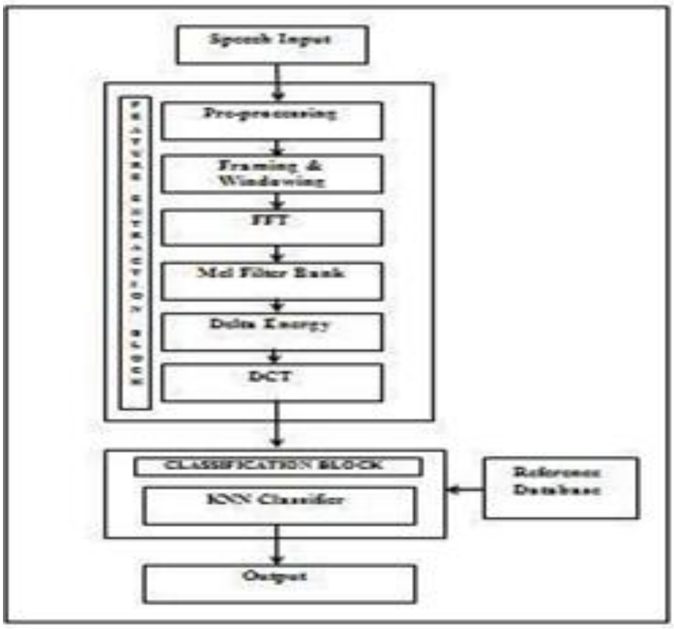

Fig.1: Schematic Representation Of Speech Recognition System

In this system we have used raspberry pi 3 . MFCC method is used for feature extraction of the voice samples and for the classification purpose the KNN classifier is used. With the specific end goal to extract the features, the voice sample is taken as an information as input. After that the features are extracted from that speech signal of each sample. Then those features given to the classifier which is KNN classifier for the classification purpose. In this KNN classifier the Euclidian distance function is used to find the particular class which one is its K-NN which is measured by a distance function. After getting the k nearest neighbor distance the resulting output get displayed.

A. Speech Recognition

In previous studies, automatic speech recognition depends on the hybrid or measurable model, for example, Hidden Markov Models. These models are prepared on several hours in which the information recorded by numerous speakers. The automatic verbalization apperception system is not suitable for the speakers with highly verbalization disorders because of the material is variable factor and it is excessively not quite the same as would be expected discourse utilized as a part of the preparation models. For speaker dependent recognition in which the WRA decreases with expanding the huge vocabulary size [17].This minimization is exponentially increases when the input speech is highly variable. So in the training phase, the speech recognition systems can be re-created utilizing both the using both the training or preparation samples as well as the collected samples with the user preparing application which produces recognizer with high accuracy and variety in the speaker’s speech. And then this procedure can be rehashed.B. FEATURE EXTRACTION (MFCC):

Feature extraction is a procedure that extracts a little measure of data or information from the voice signal which can be utilized to represent every verbalizer Feature extraction intends to convert a signal into a kind of parametric representation for further investigation and processing. The verbalization signal is a gradually or gradually time varying signal. MFCCs predicated on critical bandwidth of human auditory perceiver with frequency, it gets distributed straightly at low and logarithmically at high frequencies which gets utilized to capture consequential characteristics. A frame size of 20 milliseconds is utilized for the feature extraction from the verbalization signal.

1. Frame Blocking: The perpetuate verbalization signal is divided into number of frames i.e. N samples with the adjacent frames which are disunited by M. ($M<N$).
2. Windowing: The window applied to those frames to minimize the discontinuities between the frames. If the window is defined as $w(n), 0<n\leq N-1$, then the result of windowing is,

$yl(n)=x1(n)w(n);\ 0\leq n\leq N-1$.......(1)

For hamming window,

$\omega(n) = 0.54 - 0.46\cos[\frac{2\pi n}{N-1}]$....(2)

3. FFT: The FFT convert each frame of sub samples into frequency domain. FFT minimizes the computation time and as well as complex multiplications. FFT is an expeditious algorithm which is utilized to implement the DFT as,

$$Xn = \sum_{k=0}^{N-1} xk\, e^{-2\pi jkn/N}, n = 0,1,2,\ldots,N-1$$
…(3)

4. Mel-frequency Wrapping: The human recognition of frequency contents of sunds for verbalization signal does not follow a linear scale. Formula to calculate mels as below,

mel(f) = 2595*log10(1+ f / 700)……..(4)

The filterbank is applied in frequency domain. The width of the triangular filter fluctuates so the center frequency is included in the critical band.

5. Cepstrum: Cepstrum is the FT of logarithm of autospectrum.. The numbers of coefficients are obtained after wrapping. At long last , the (IDFT) is used to estimate the cepstral coefficients. So, it changes log domain coefficients into frequency domain. The result is called as MFCC. By using (DCT) the mel cepstrum coefficients gets converted into time domain. Formula to calculate MFCC as,

$$C_n=\sum_{k=1}^{K}(\log S_k)\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right], \quad n=1,2,\ldots\ldots,K$$
……(5)

C. **Classification (K-NN Classifier)**:

K-NN is an easiest classification model that show that stores every single accessible case and characterizes new cases in view of a likeness measure (e.g., separate capacities). K-NN is likewise utilized as a part of the statistical estimation. It is supervised learning calculation algorithm. It characterizes the new information such as input speech signal which is based on the nearest neighbor classifier. It figure least distance between the input information as a speech signal and the training samples. The K-NN decided lager part voting of the closest neighbor class for the information signal. The input information speech sample will compare with all the training speech samples and then the K-NN decides the high response time. In this study, for every one of the information signal, least separation from input information signal to the training as a preparation signal is calculated to find the classification of the training as preparation database. An Euclidean distance function is utilized to compute how every one of the training database of the voice samples to the test signals that is being analyzed. Euclidean Distance can be calculated as:

$$d_E(x,y) = \sum_{i=1}^{N}\sqrt{x_i^2 - y_i^2}$$
……….(6)

According to this the minimum distance gets measured and the resultant outputs get displayed.

D. **Raspberry Pi 3 Model B:**

The Raspberry Pi is a charge card estimate arrangement of little single-board PCs to advance the educating of essential software engineering in schools and in creating nations.

In this study we have used Raspberry pi 3 model B.RAM of 1GB which is remains same as Raspberry pi 2 as wellas graphics capabilities, provided by the Video Core IV GPU.We have installed some libraries in Raspberry Pi which are usedin the programming of feature extraction and classification.Libraries such as scipy, numpy, etc which are used formathematical calculations, etc. Raspberry Pi 3 inbuilt having on-board Wi-Fi and Bluetooth. Wi-Fi, wireless keyboards, andwireless mice. In this study Python as a programming languageis used for feature extraction as well as classification purpose.This system is used in multiple systems such as in medicalapplications, security monitoring, health care, home automation,etc.

.

E. Database:
We have used speech of the severe speech disordered speakers. We have recorded 26 alphabets of each of the speaker.
We have used total 390 samples. A class of 15 samples each.

IV. Results

1. Confusion Matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	10	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
B	0	11	0	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
C	0	0	10	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0
D	0	0	0	11	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
E	0	2	0	1	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
F	2	0	0	0	0	9	0	2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
G	0	0	2	0	1	0	10	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	3	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
I	1	0	0	0	0	0	0	0	10	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
J	2	0	0	0	0	0	0	0	0	10	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	2	0	0	0	0	0	0	0	0	3	9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
L	0	0	0	0	0	0	0	2	0	0	0	11	0	2	0	0	0	0	0	0	0	0	0	0	0	0
M	1	0	0	0	0	2	0	0	0	0	0	2	10	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	2	0	2	11	0	0	0	0	0	0	0	0	0	0	0	0
O	1	0	0	0	0	0	0	0	2	0	0	0	0	0	10	0	0	0	0	0	2	0	0	0	0	0
P	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	9	0	0	0	3	0	0	0	0	0
R	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	2	0
S	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0
T	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0
U	0	0	0	0	1	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	8	0	3	0	0	0
V	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	10	0	0	0	0
W	0	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	8	0	0	0
X	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	9	0	0
Y	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	10	0
Z	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	10

Table No.: Confusion Matrix

- Accuracy = (TP + TN)/(P+N) Total Accuracy is gives as:

A	95.38%	N	98.46%
B	96.92%	O	98.20%
C	97.17%	P	98.46%
D	97.94%	Q	98.46%
E	95.64%	R	98.71%
F	97.43%	S	96.92%
G	98.46%	T	98.20%
H	96.66%	U	96.41%
I	96.66%	V	97.94%
J	96.66%	W	97.43%
K	95.64%	X	96.92%
L	98.46%	Y	97.69%
M	98.20%	Z	98.71%

Table No.2: Accuracy Of Every Alphabet

Total Accuracy= 97.45%

We achieved better accuracy as compared to the accuracy of previous studies. We have used 390 speech samples which contain normal and abnormal speech in the form of alphabets.

2. Cross Correlation Between Normal & Abnormal Alphabet U:

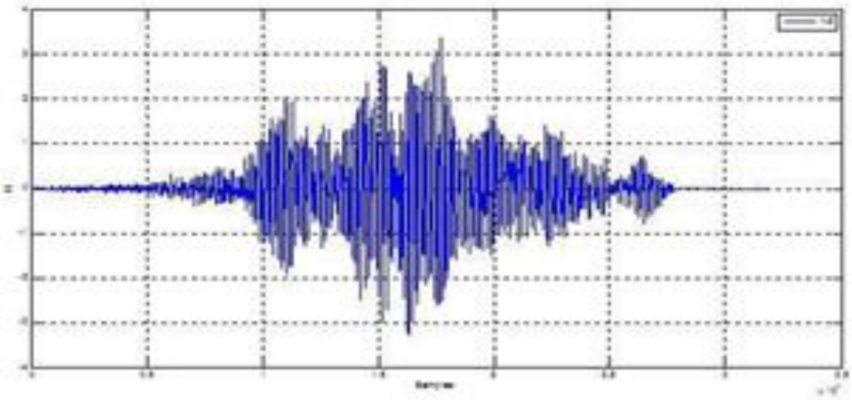

Fig.:2: Cross Correlation Signal

3. Power Spectrum Of Alphabet H:

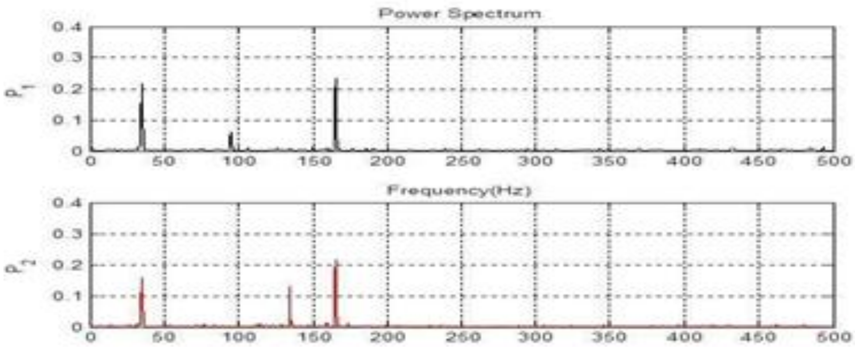

Fig:3: Power Spectrum

4. Spectrogram Plot Of Normal & Abnormal Aphabet P:

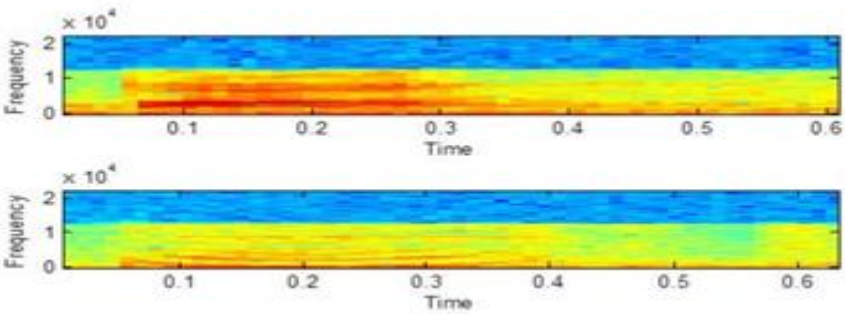

Fig.:4: Spectrogram Plot

5. Final Experimental Results:

In this Study we get the following results such as Figure. 5 shows the input speech signal which is given as an input to the system. Figure 6 shows the extracted signal of input speech signal. After that it calculates the minimum distance to the training database by using Euclidean Distance function in the K-NN classifier. Figure 7 shows the resulting output in alphabet.

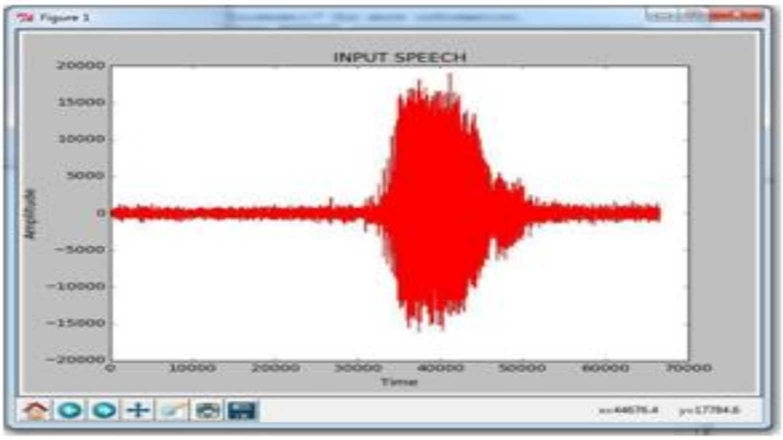

Fig. 5. Input Speech Signal

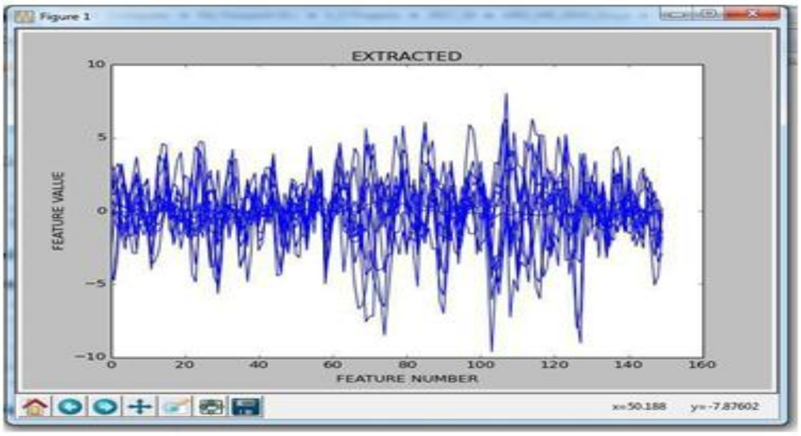

Fig. 6. Extracted Signal

Fig. 7. Resulting Output

V. Conclusion And Future Scope

This system has described the development of the speech recognition system which is an acceptable system for the people with severe speech disabilities. It able to construct recognizer with apperception precision in excess of 85%.The word disability can avert those people who are suffering from severe speech disabilities from communicating in a way of doing thing that allows them to use for one ends their potential in education recreation. It ready to build recognizer with acknowledgment precision more than 95%.

A portion of a few issues which restrict the execution of the framework. Those constraints will be center in future work.

References

[1]. P. Enderby and L. Emerson, Does Speech and Language Therapy Work?. London, U.K.: Singular, 1995.
[2]. C. L. Kleinke, “Gaze and eye contact: A research review,”Psychol. Bull., vol. 100, no. 1, pp. 78–100.
[3]. B. O’Keefe, N. Kozak, and R. Schuller, “Research priorities in augmentative and alternative communication as identified by people who use AAC and their facilitators,” AugmentativeAlternative Commun., vol. 23, no. 1, pp. 89–96, 2007.
[4]. L. J. Ferrier, H. C. Shane, H. F. Ballard, T. Carpenter, and A.Benoit, “Dysarthric speakers’ intelligibility and speech characteristics in relation to computer speech recognition,”Augmentative Alternative Commun., vol. 11, no. 3, pp. 165– 175, 1995.
[5]. N. Thomas-Stonell,A.L.Kotler, H.A. Leeper, and P. C.Doyle, “Computerized speech recognition: Influence of intelligibility and perceptual consistency on recognition accuracy,” Augmentative Alternative Commun., vol. 14, no. 1, pp. 51–56, 1998.
[6]. P. D. Polur and G. E. Miller, “Experiments with fast Fourier transform, linear predictive and cepstral coefficients in dysarthric speech recognition algorithms using hidden markov model,” Neural Systems and Rehabilitation Engineering, IEEETransactions on, vol. 13, no. 4, pp. 558–561, 2005.
[7]. M. S. Hawley et al., “A speech-controlled environmental control system for people with severe dysarthria,” Med. Eng. Phys., vol. 29, no. 5, pp. 586–593, 2007.
[8]. P. D. Polur and G. E. Miller, “Investigation of anHMM/ANN hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals,”Medical engineering & physics, vol. 28, no. 8, pp. 741–748, 2006.
[9]. Mark S. Hawley, Stuart P. Cunningham, Phil D. Green, PamEnderby, Rebecca Palmer, Siddharth Sehgal, and Peter O’Neill “A Voice-Input Voice-Output Communication Aid for PeopleWith Severe Speech Impairment” IEEE transactions on neural systems and rehabilitation engineering, vol. 21, no. 1, january 2013.
[10]. “An Overview of Technical Progress in Speech Recognition”, International Journal of Advanced Research inComputer Science and Software Engineering, Volume 3, Issue 3, March 2013.
[11]. “Speech Recognition of Articulatory Handicapped People By Using LPC” International Journal Of Scientific ResearchAnd Education,Volume3,Issue 2,Pages-2893-2899, February-2015.
[12]. “Effect of Pre-processing along with MFCC Parameters inSpeech Recognition”, 2016 IJEDR | Volume 4, Issue 3 | ISSN:2321-9939.
[13]. Representation Learning based Speech Assistive System for Persons with Dysarthria Citation information: DOI 10.1109/TNSRE.2016.2638830, IEEE Transactions on Neural Systems and Rehabilitation Engineering.
[14]. Implementation of Speech to Text Conversion International Journal of Innovative Research in Science, Engineering (An ISO 3297: 2007 Certified Org.) July 2015.
[15]. R. Palmer, P. Enderby, and S. P. Cunningham, “The effect of three practice conditions on the consistency of chronic dysarthric speech,” J. Med. Speech-Language Pathol., vol. 12, no. 4, pp. 183–188, 2004.
[16]. M. Parker, S. P. Cunningham, P. Enderby, M. S. Hawley, and P. D. Green, “Automatic speech recognition and training for severely dysarthric users of assistive technology: TheSTARDUST project,” Clin. Linguistics Phonetics, vol. 20, no.2–3, pp. 149–156, 2006.
[17]. “MFCC based recognition of repetitions and prolongations in stuttered speech using K-NN and LDA.” ,Proceesing of 2009 IEEE student conference on researdch & development(SCOReD 2009), 16-18 Nov.2009, UPM serdang, Malaysia.

Prof. S.S. Bhabad "Communication Aid For People With Severe Speech Disability "IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) , vol. 8, no. 5, 2018, pp. 01-08