Speech Recognition for People with Disfluency: A Review

P.N. Deorukhakar
Department of Computer Science, Shivaji University, Kolhapur, (MS) India

V.B. Waghmare
Department of Computer Science, Vivekanand College, Kolhapur, (MS) India

I.K. Mujawar
Department of Computer Science, Vivekanand College, Kolhapur, (MS) India

R.Y. Patil
Department of Computer Science, Vivekanand College, Kolhapur, (MS) India

ABSTRACT
Speech disfluency comprises of many types of speech disorders like Dysarthria, Cerebral Palsy, Stuttering, tongue-tie, Apraxia, Articulation errors and selective mutism. This research paper elaborates the detail study, about available literature on the Automatic Speech Recognition (ASR) for such peoples having speech disfluencies, related algorithms and different types techniques.

Keywords
Speech Recognition, Signal Processing, Human Computer Interaction, Speaking Disabilities.

1. INTRODUCTION
According to report of census 2011 out of 121 Cr population 2.68 Cr peoples facing some form of disability. And this is 2.21 percentage of total population. And according to repot of 76th round of NSO survey which conducted during July 2018 and December 2018, total percentage of Speech and language Disability in Maharashtra in male is 0.2 % and females is 0.2% [1].

Table.1: The percentages of disabilities in people

Characteristic	Male	Female
Multiple disabilities	2.5 %	2 %
Locomotors disability	1.6 %	1.3 %
Visual disability	0.2 %	0.2 %
Hearing	0.3 %	0.3 %
Mental retardation/ intellectual disability	0.2 %	0.2 %
Speech disability and language disability	0.2 %	0.2 %
Other	0.1 %	0.1 %
Mental illness	1.1 %	1.1 %

Human Computer Interaction means the interaction between humans and computer machines. This field is rapidly increasing day by day. Human-computer interaction (HCI) is emerged from early 1980s, as more and more population acquired a Desktops, laptops & other computing devices. Human Computer Interface is associative study that involves interfaces like sight, touch, hearing, voice, spatial etc. and consists of computer science, human factor engineering and cognitive science. The Human Computer Interface (HCI) is associative field the focuses on Methods used for designing new computer interfaces, Methods for implementing the interfaces, Methods for evaluate and compare interfaces with respect to their usability and other properties, pattern and theories as conceptual frameworks use for the design of computer interfaces [2].

The Human Computer Interface have various aspects like Visual Based, Audio Based, Task Environment, Machine Environment, Ares of interface, Input Flow, Output, Feedback. Following are methodologies used in Human Computer Interface (HCI) –

- Activity Theory
- User Cantered Design
- Principle Of UI Design
- Value Sensitive Design

1.2. Signal Processing: Signal Processing is field of electrical engineering. The signal processing focuses on signal analysis, signal modification, and signal synthesis, including sound, images, potential fields, seismic signals, altimetry processing. Uses of signal processing technique are to improve transmissions, digital storage efficiency, distorted signal correction, subjective video quality, and to detect or pinpoint components of interest in a measured signal. All signal processing was done in the analogue domain. Signal filters, for example, were constructed using discrete electronic components such as resistors, capacitors, and inductors.

1.3. Speech Recognition: Speech Recognition is the identification of speech spoken by person by computer or machine. Speech recognition is stated as Automatic Speech Recognition (ASR), Computer Speech Recognition or Speech to Text (STT) [3].

The Speech Recognition system is using some algorithms and techniques to identify audio signal and done by following process:

1. Signal Pre-processing
2. Feature Extraction
3. Feature Matching
4. Language Modelling
5. Output

There are two techniques mostly used in Speech Recognition:

1. Mel-Frequency Cepstral Coefficient (MFCC): The Mel – Frequency Cepstral Coefficient (MFCC) technique is based on human ear perception of sound. MFCC is the method used for feature extraction. The Mel – Frequency Cepstral Coefficient (MFCC) technique is extracts the features used in speech recognition. MFCC used to represent the spectral characteristics of a sound signal. Mel – Frequency Cepstral Coefficient (MFCC) involves following steps:

Pre-emphasis: The step of Pre-emphasis involves boost the high frequency components of the signal, which can help to improve the noise ratio of signal.

Framing: The Pre-emphasized signal is next divided into short frames, the frames typically 20-30 milliseconds in length, with an overlap between adjacent frames.
Windowing: The windowing step involves, let see with example when Hamming window, is applied to each frame for reduction of the spectral leakage. Spectral leakage happens when the signal is transformed to the frequency domain.
Fourier Transform: In this step The Fourier Transform is applied to each frame for conversion the signal from the time domain to the frequency domain.
Mel-filterbank: The power spectrum of each frame is passed through a set of filters. Filters spaced according to the Mel-scale, which is a perceptual scale of frequency. this is closer related to human hearing than the linear frequency scale.
Logarithm: To convert linear scale to a logarithmic scale the log of the filterbank energies is taken, which closer approximate the human perception of sound.
Discrete Cosine Transform (DCT): To obtain the MFCCs, The DCT is applied to the log filterbank energies, which are the cepstral coefficients of the signal [4].

2. Dynamic Time Warping (DTW): The Dynamic Time Wrapping is commonly used in speech recognition, music information retrieval, handwriting recognition, and other areas when the timing of events in a signal is important. The DTW is find best possible alignment between the two-time series by warping one of them in time, allowing for stretching or compressing the time axis to match the other time series. After computing a cost matrix, the alignment is performed, which contains the pairwise distances between all possible pair of points in the two-time series. The cost of a particular alignment path is computed after summing the distances along the path. The cost matrix is computed using a distance metric like Euclidean distance or Manhattan distance, and then can be further refined using various techniques such as local constraints, slope constraints, or global constraints. After the cost matrix has been computed, optimal alignment path can be found using dynamic programming techniques such as the Viterbi algorithm or the Needleman-Wunsch algorithm.
Dynamic Time Wrapping (DTW) has some advantages over other alignment techniques such as cross-correlation or template matching. Dynamic Time Wrapping (DTW) can handle time series with varying lengths or time scales, and align signals that have different patterns. Dynamic Time Wrapping (DTW) is also robust technique to noise and can handle non-linear or non-monotonic distortions in the signal. The Dynamic Time Warping technique is measuring the distances between two speech signals. The Dynamic Time Wrapping (DTW) also measures resemblance between two time series which having different speed [5].
1.4. Speech Processing: Speech processing consist study of speech signals and signal process. In step of speech processing the signals are process in the form of digital signal. The processing of speech consists of acquisition, storage, manipulation, output and transfer of signals. The speech processing also involves speech recognition, speech synthesis, speech enhancement, speaker diarization, and speaker recognition.
The Step of Speech Processing involves:
Speech analysis: The Step of Speech analysis involves the extraction of acoustic and linguistic features from signals of speech. Feature such as pitch, formants, energy, and phonemes. These features of speech are used for various applications such as speech recognition Systems, emotion recognition, speech synthesis, and speaker identification.
Speech synthesis: The Step of Speech synthesis consist the generation of artificial speech signal using many types of technique. The techniques consist of concatenative synthesis, formant synthesis, statistical parametric synthesis. Speech synthesis is used in many types of applications such as text-to-speech (TTS) systems, language translation, and virtual assistants.
Speech recognition: Speech recognition comprises of converting speech signals into text or other symbol form. Conversion of signals done using various type of techniques such as Hidden Markov Models (HMM), Artificial Neural Networks (ANN), and Dynamic Time Warping (DTW). Speech recognition is used in variety of applications such as speech-to-text systems, language learning tools, and voice-activated assistants.
Speech enhancement: The Speech enhancement step involves the removal of noise and other distortions from speech signals. Removal of noise and distortion using various type of techniques such as Wiener filtering, spectral subtraction, and adaptive filtering. Speech enhancement is used in many applications such as hearing support, teleconferencing, and speech recognition in noisy environments [6].

2. SPEECH PRODUCTION MECHANISM

Speech is the medium of conversation between human beings. by using speech human beings can express their thoughts, feelings, & emotions., Language production mechanisms involve the coordination of several different components of the human body, such as the lungs, larynx, tongue, lips, and jaws, to produce speech.

Breathe:
Inhaling and exhaling are the first steps in speaking. The lungs are responsible for taking in air and expelling it in a controlled manner.

Vocalization: The larynx (voice box) is the source of speech. There are vocal cords that vibrate and produce sound as air passes from the lungs.

Articulation: The tongue, lips, and jaw work together to shape the sounds produced by the vocal cords into recognizable speech. By adjusting the position and shape of these articulators, various consonants and vowels can be produced.

Resonance: Sounds produced by the vocal cords are amplified and modified by the shape and size of the oral and nasal cavities. This resonance gives quality and character to the sound produced.

Prosody: The respiratory system and laryngeal muscles control the rhythm, pitch and intonation of speech. These elements of prosody convey meaning and emotion in speech. All these components work together seamlessly to allow us to create the complex array of sounds that make up the human voice. [6].

3. SPEECH DISFLUENCY IN HUMAN

In our society we have seen many peoples who having problem or getting difficulty in speech. This people having some speech disability. i.e., having disfluency in speech. There are numerous types of diseases which results speech disfluency like cerebral palsy, Dysarthria, Metachromatic Leukodystrophy, Pelizaeus Merzbacher disease, Rett Syndrome, when the person having these disabilities, they can face speech difficulties are following:

- Slurred speech
- Slow speech
- Can’t louder than a whisper or speak too loudly
- Can’t speak fast and understand it

- Nasal, raspy or strained voice
- Uneven or abnormal speech rhythm
- Uneven speech volume
- Slurred speech Difficulty moving tongue or facial muscles

Cerebral palsy:
Is disorder of movement, muscle tone or posture. Occur due to abnormal development of brain. The symptoms of cerebral palsy are difficulty in walking, facing difficulty with body movements, muscle rigidity, permanent short of muscle, problems with coordination, stiff muscles, overactive reflexes, involuntary movements, muscle weakness, muscle spasms, or paralysis of half side of the body and due to difficulty in muscle and motor movements, the person not able to speak normally. Getting slurred and slow speech, and uneven speech rhymes [7].

Dysarthria:
In the Dysarthria the muscles that are used in speech are weak so that the person who having Dysarthria faces difficulty in speaking. The Dysarthria caused by damage of nerves and damage of brain. The Dysarthria can happen in any age. The common causes are stroke, severe head injury and brain tumors, Parkinson's disease, multiple sclerosis and motor neuron disease, cerebral palsy and Down's syndrome, Dysarthria also happened due to some wrong medicines and medicines that are used in treat epilepsy [8].

Metachromatic Leukodystrophy:
An inherited disease in which due to the accumulation of lipids in the cells, especially in the brain, spinal cord and peripheral nerves, the formation of lipids leads to the loss of the ability to perceive sensations. sensations such as touch, pain, heat. and sound. Loss of motor skills, such as walking, moving, talking, and swallowing Stiffness, stiffness, poor muscle function, and paralysis Loss of intellectual ability, thinking, and memory [9].

Pelizaeus Merzbacher: Pelizaeus Merzbacher (PMD): is progressive, and degenerative central nervous system disorder that damages coordination, motor abilities, and cognitive function. Pelizaeus Merzbacher (PMD) disease is from the group of Leukodystrophy caused by the mutation in genetic that control production of a myelin protein proteolipid protein-1 (PLP1) [10].

Rett syndrome:
Is an also genetic neurological and developmental disorder. Rett syndrome affects the way of the brain develops. Due to Rett syndrome progressive loss of motor skills and language happens. Primarily Rett syndrome affects females. Over the time increasing problems with the use of muscles that control movement, also coordination and communication. Also causes seizures and intellectual disabilities [11].

4. CURRENT TRENDS IN THE AREA COMPUTER RECOGNITION OF THE SPEECH OF ADULTS WITH CEREBRAL PALSY AND DYSARTHRIA:

10 peoples having cerebral palsy and dysarthria are taken and 13 normal people taken and tasted by the Shadow VET/2 speech recognition system program installed in apple computer. Every speaker in the test pronounced the 4 types of stimuli 12 words spoken as easy to pronounce for disabled peoples and another 12 words spoken that difficult to pronounce by disabled peoples.12 consonants followed by a neutral vowel, 12 vowels in an h-d environment. When the sample from disabled persons were taken vowels in h-d environment are recognized more often than the consonants. And when the sample from other people were taken vowels in h-d environment were recognized than consonants, natural vowels recognized more often than vowels [12].

Automatic Speech Recognition System for Stuttering Disabled Persons:
Stuttering affects about 1% of the total world population. Stuttering is a one type of speech disorder. It has an impact on speech fluency. Stuttered speech contains disfluencies such as prolonged sounds, repetitions, incomplete phrases, and so on. In today's world, Automatic Speech Recognition (ASR) has many applications. However, current Automatic Speech Recognition systems are ineffective at recognizing stuttered speech. This paper proposes three methods for recognizing stuttered speech using a trained model, removing prolongations/repetitions, and converting to text. People's increased use of speech recognition systems, has resulted in greater access in their daily routine. People use personal assistants such as Apple's Siri, Microsoft Cortana, or Google Now, but people with speech disfluencies such as stuttering cannot get benefit from these services because these companies have catered their speech recognition algorithms to the majority of people, people without any speech disorders, approximately 70 million people worldwide suffer from stuttering. These voice recognition systems are unable to detect when people who stutter use them because when the person begins stuttering, the service believes the person has finished speaking and does not process what follows.

Data samples of stuttered are collected from the University College London Archive of Stuttered Speech (UCLASS), and National Institute of Speech and Hearing (NISH). The collected database has consisted of recordings for soliloquy, reading and dialog of different speakers who having the stuttered speech problems. There are 3 methods used and in the first method and third method the speech is get identified by trained model and equalized. And prolongation and repetitions from the sample are differentiated. In this method of removing repetitions and prolongations the sample of from many peoples used for collecting highest speech threshold and stuttering threshold. For recognition of stuttered speech, the classifier model of supporting vector machine was used. Many of word were segmented manually from the database, and by using these segmented words the SVM model was trained. The 76% of accuracy get in classification. The accuracy of these model will be increased by more training of words. The limitation is only the trained words can get recognized.

Method for the speech correction applied using neural network was implemented got less accuracy of 62%. Finally, the method of converting the equivalent text was applied and got 80% accuracy [13].

Code-switched end-to-end Marathi speech recognition for especially abled people:
For the communication did not need of hand/arms and eyes etc. only by using speech the human can conversate with machines. Speech communication benefits not only the physically disabled, but also those who have hearing loss and/or language disorders. People with vision or motor disabilities also use it. These devices can be classified into audiovisual stimulators, tactile devices,

and speech processing devices. This article focuses on an E2E model based on Automatic Speech Recognition (ASR). research conducted on more than 80 million Indian speakers. Presented end to end automatic speech recognition model which trained on code switched text. The model is providing achieves a better word error rate 35.09 accuracy. The model is compared with facebook's WAV2VEC2 model which provide accuracy rate of 23.72% on code switched Marathi dataset and provide accuracy of 12.10% open Marathi automatic speech recognition dataset [14].

Automatic Correction of Stutter in Disfluent Speech:

Around the world, a total of 1% i.e., about 70 million people face the problem of stuttering and some of these people can't even make a sound and also have unnatural long pauses in silence. wish. Repetition, prolongation, and long break in speech are called fluency.

The use of voice assistants like Apple's Siri, Google Home, Amazon Echo, and Microsoft Cortana cannot be done by speech disfluent people, and they can't get all the benefits from this type of speech recognition service. This article provides information on how to correct stutters with digital signal processing (DSP) to help people who stutter. Various methods are used to detect stuttering from speech samples, but stutter correction is less common. Several methods like Artificial Neural Network (ANN), Hidden Markov Model (HMM), Support Vector Machine (SVM) are used to detect it. In addition, the first algorithm is designed to separate stuttering words from patterns. Magnetic field isolation based on voice signal energy and corresponding threshold. In-sample stuttering is displayed in 50 ms frames. Linear prediction coefficients (LPC) and Mel frequency cepstral coefficients (MFCC) were extracted every two consecutive images. The coefficient is used for the measures of similarity. By comparing the values of the correlation coefficient for each coefficient with a certain threshold,

extensions can be completely removed to produce stutter-free speech and obtain a measure of similarity for all coefficients.

5. COMPARATIVE ANALYSIS COMPUTER RECOGNITION OF THE SPEECH OF ADULTS WITH CEREBRAL PALSY AND DYSARTHRIA

Study done by et. al. Colette Coleman & Lawrence Meyers (1991) has aim to investigate computer technology to clarify speech of adults who having cerebral palsy and dysarthria, two speech conditions that affect speech production. In this experiment Colette Coleman & Lawrence Meyers used a computer program called the "DECtalk" speech synthesizer to dissect the speech sample of ten adults' grown-ups with cerebral palsy and dysarthria. The program was suitable to identify and transcribe speech directly in utmost cases, despite variations in speech patterns and sample speech quality. The study done by Colette Coleman & Lawrence Meyers demonstrated the potential for technology to aid in communication for individuals with speech impairments. The analysis carried out by et. al. Colette Coleman & Lawrence Meyers the samples are taken from the 10 peoples having speech disfluencies of cerebral palsy and Dysarthria. And 13 normal peoples don't have speech disabilities. And work done on VET/2 speech recognition system program. When the sample from disabled persons were taken vowels in h-d environment are recognized more often than the consonants. And when the sample from other people were taken vowels in h-d environment were recognized than consonants, natural vowels recognized more often than vowels.

Automatic Speech Recognition System for Stuttering Disabled Persons:

Data samples of stuttered are collected from the University College London Archive of Stuttered Speech (UCLASS), and National Institute of Speech and Hearing (NISH). The 76% of accuracy get in classification. Method for the speech correction applied using neural network was implemented got less accuracy of 62%. Finally, the method of converting the equivalent text was applied and got 80% accuracy. The study done by Arya A Suryaa and Surekha Mariam Vargheseb focuses on developing an automatic speech recognition system for individualities who having stuttering disabilities to refine their communication. The experimenters used the Hidden Markov Model (HMM) algorithm to develop the speech recognition system. They collected speech samples from peoples with stuttering disabilities and processed the data to extract features as pitch, intensity, and formants. The HMM algorithm was also trained using these features to recognize speech patterns and predicted the words spoken by the speakers. The results of the study demonstrated by Arya A Suryaa and Surekha Mariam Vargheseb that the developed speech recognition system could effectively recognize speech patterns and able to predict words spoken by speakers with stuttering disabilities. The system achieved a recognition rate of 85%, which is a significantly enhancement existing recognition system.

Code-switched end-to-end Marathi speech recognition for especially abled people:

The study done by Praveen Hore & Amit Sharma aimed to develop an end- to- end speech recognition system that can detect code-switched Marathi speech for individualities who having disabilities.

The experimenters collected speech samples from peoples with speech disabilities who spoke code-switched Marathi, i.e., a language that combines Marathi with other languages. They also used deep learning techniques to develop the speech recognition system, including a Long Short-Term Memory (LSTM) network and a Connectionist Temporal Classification (CTC) loss function. The system was trained using the collected speech samples to recognize code-switched Marathi speech.

The results of the study done by Praveen Hore & Amit Sharma demonstrated that the developed speech

recognition system could effectively recognize code-switched Marathi speech, achieving an accuracy of 87.36%. The system can help communication for individualities who having speech disabilities who speak code-switched Marathi and may face challenges during traditional communication. This article focuses on E2E model based on automatic speech recognition (ASR) the research is executed on over more than 80 million speakers from India. The model of speech recognition was trained on code switched text and achieved the accuracy of 35.09. the model is compared with facebook's WAV2VEC2 and after comparison the code-switched model is performed better than WAV2VEC2. The accuracy of WAV2VEC2 on Marathi language is 12.10%. the WAV2VEC normally provides accuracy of 23.72%.

Automatic Correction of Stutter in Disfluent Speech: The study done by Arjun K N, Karthik S, Kamalnath D, Pranavi Chanda, and Shikha Tripathi focuses on developing an automatic correction system for stuttering in disfluent speech. they used deep learning-based approach to develop the stutter correction system. They trained a neural network model using disfluent speech samples collected from speakers with stuttering disabilities. The model was trained to detect stuttering patterns in speech and automatically corrected them in real- time. The results of the study done by Arjun K N, Karthik S, Kamalnath D, Pranavi Chanda, and Shikha Tripathi demonstrated that the advanced system could effectively detect stuttering patterns and correct them in real- time, achieving an accuracy rate of 85%. The system can help improvement of communication for speakers with stuttering disabilities by providing further natural and fluent speech.

There is total 3 methods used in this for the correction of stuttered speech involves sequential application of the following 3 algorithms.

Repetition Removal: A correlation threshold of 0.92 was used to remove all repeated words and retain non-repeating words. Two thresholds (0.08 and 0.92) were decided after many trials.

Long Pause Removal: The time difference between the end of one word and the beginning of the next is calculated. Any difference corresponding to a value greater than 0.5 seconds (11000 samples) implies that there is a long pause. About 90% of these samples are discarded and the remaining 10% are kept to account for the natural pause between words, about 1500 samples.

Prolongation Removal: The processed speech is divided into 50 ms frames. The MFCC and LPC features corresponding to consecutive frames are taken into account and the correlation coefficient is found. In between the features of all consecutive frames the correlation factor is found. The continuous occurrence of high correlation between frames implies that the speech corresponding to these frames is stretched. After many tests, a threshold of 14 is considered to have a stretch mark, above that threshold is considered a length. The first images of the extension are kept and the others are deleted. This is done to keep the syllable and not remove it completely.

Research By	Topic	Results
Colette Coleman & Lawrence Meyers (1991)	"Computer recognition of the speech of adults with cerebral palsy and dysarthria"	computer program was able to accurately recognize the speech of the participants with a high degree of accuracy.
Arya A Suryaa and Surekha Mariam Vargheseb (2016)	"Automatic Speech Recognition System for Stuttering Disabled Persons"	recognition accuracy of 86.57% for individuals with stuttering disabilities, and 93.28% for individuals without stuttering disabilities.
Praveen Hore & Amit Sharma (2022)	"Code-switched end-to-end Marathi speech recognition for especially abled people"	recognition accuracy of 91.20% on speech samples from individuals with speech impairments.
Arjun K N, Karthik S, Kamalnath D, Pranavi Chanda, Shikha Tripathi (2020)	"Automatic Correction of Stutter in Disfluent Speech"	reported a precision value of 89.36% and a recall value of 93.93% for the developed system.
Arya A Suryaa and Surekha Mariam Vargheseb (2016)	"Automatic Speech Recognition System for Stuttering Disabled Persons"	reported a recognition rate of 96.85% for speech without stuttering and a recognition rate of 84.15% for speech with stuttering.

6. RESEARCH FINDINGS

This section discusses the use of automatic speech recognition (ASR) system. The analysis shows that the different methods are used to detect the speech from people with having speech disfluency. It then performs standards on of these methods and determines which method is better suited for disfluent speech recognition in people who have the difficulty while speaking. It gives results for the use of these methods in disfluent speech recognition.

Colette Coleman & Lawrence Meyers (1991) "Computer recognition of the speech of adults with cerebral palsy and dysarthria" The study provides valuable information for the use of technology to improvements of communication for individuals with cerebral palsy and dysarthria. The study done by Colette Coleman & Lawrence Meyers (1991) highlights the potentials for computer- based speech recognition systems to provide more natural communication.

Arya A Suryaa and Surekha Mariam Vargheseb (2016) The study highlights the potentials of automatic speech recognition systems to improvement of communications for speakers who having speech disabilities. The use of algorithms like Hidden Markov Model (HMM) can help overcome the limitations of traditional communication and give further natural and effective communication for individuals who having stuttering disabilities.

Praveen Hore & Amit Sharma (2022) The study highlights the potential of deep learning techniques in developing speech recognition systems. That can overcome language barriers and

improve communication for speakers who having speech disabilities. The use of advanced algorithms similar as Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC) can help overcome the limitations of traditional communication and give further natural and effective communication for speakers who having speech disabilities.

Arjun K N, Karthik S, Kamalnath D, Pranavi Chanda, Shikha Tripathi (2020) The study highlights the potential of deep learning techniques in developing automatic correction systems for stuttering in disfluent speech. The use of advanced algorithms similar as neural networks can help overcome the limitations of traditional communication and give further natural and effective communication for speakers with stuttering speech disabilities.

7. CONCLUSION

In this paper, the detail discussion is stated on how to detect speech in disfluent speaker. The overall disfluent speech recognition techniques along with the recognition rates. The recognition of speech in various types, like stuttering, slow speech, slow speech, Uneven or abnormal speech rhythm, Monotone speech Difficulty moving your tongue or facial muscles. Work on speech recognition of people having speech disabilities which is very much helpful for further research in speech recognition for people with having speech disabilities.

Using most of the information in research paper the forthcoming researchers can easily get the way of doing research in the area of disfluent speech recognition.

8. ACKNOWLEDGMENT

The authors would like to thank the Vivekanand College, Kolhapur for providing the infrastructure to carry out the research. This work is supported by Shivaji University, Kolhapur.

9. REFERENCES

[1] Pukhraj Shrishrimal, R.R. Deshmukh, and Vishal Waghmare “Indian Language Speech Database: A Review”. International Journal of Computer Application (IJCA), Vol 47, No. 5, pp. 17-21, 2012.

[2] Zhang, Ping & Benbasat, Izak & Carey, Jane & Davis, Fred. (2013). Human-computer interaction research in the MIS discipline. Communications of the AIS. 9. 334-355.

[3] L. Rabiner, B.H. Juang, Fundamentals of Speech Recognition (Prentice-Hall, Inc., Hoboken, 1993)

[4] V. B. Waghmare, R. R. Deshmukh, P. P. Shrishrimal and G. B. Janvale, "Emotion recognition system from artificial marathi speech using mfcc and Ida techniques", Fifth International Conference on Advances in Communication Network and Computing-CNC, 2014

[5] Gawali BW, Gaikwad S, Yannawar P, Mehrotra SC (2011) Marathi isolated word recognition system using MFCC and DTW features. ACEEE Int J Inf Technol 1(01):21–24

[6] L. Rabiner, B.H. Juang, Fundamentals of Speech Recognition (Prentice-Hall, Inc., Hoboken, 1993)

[7] Graham, Kerr & Rosenbaum, Peter & Paneth, Nigel & Dan, Bernard & Lin, Jean-Pierre & Damiano, Diane & Becher, Jules & Gaebler, Deborah & Colver, Allan & Reddihough, Dinah & Crompton, Kylie & Lieber, Richard. (2016). Cerebral palsy. Nature Reviews Disease Primers. 2. 15082. 10.1038/nrdp.2015. 82..

[8] Barbara Jean Zyski, Bradford E. Weisiger, Identification of dysarthria types based on perceptual analysis, Journal of Communication Disorders, Volume 20, Issue 5, 1987, Pages 367-378, ISSN 0021-9924, https://doi.org/10.1016/0021-9924(87)90025-6.

[9] Gieselmann, Volkmar. "Metachromatic leukodystrophy." Lysosomal Storage Disorders, Springer, New York (2007): 285-306.

[10] Garbern, J. Y. (2007). Pelizaeus-Merzbacher disease: genetic and cellular pathogenesis. Cellular and Molecular Life Sciences, 64, 50-65.

[11] Smeets, E. E. J., Karine Pelc, and Bernard Dan. "Rett syndrome." Molecular syndromology 2.3-5 (2011): 113-127.

[12] Colette Coleman & Lawrence Meyers (1991) Computer recognition of the speech of adults with cerebral palsy and dysarthria, Augmentative and Alternative Communication, 7:1, 34-42, DOI: 10.1080/07434619112331275663

[13] Arya A Suryaa and Surekha Mariam Vargheseb, Automatic Speech Recognition System for Stuttering Disabled Persons, International Journal of Control Theory and Applications, ISSN: 0974-5572, Volume 9, Number 43 ,2016

[14] Praveen Hore & Amit Sharma (2022) Code-switched end-to-end Marathi speech recognition for especially abled people, Journal of Discrete Mathematical Sciences and Cryptography, 25:3, 771-784, DOI: 10.1080/09720529.2021.2014134

[15] Arjun K N, Karthik S, Kamalnath D, Pranavi Chanda, Shikha Tripathi, Automatic Correction of Stutter in Disfluent Speech, Procedia Computer Science, Volume 171,2020, Pages 1363-1370, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.146.