

# Automatic Speech Recognition system in Marathi for Cerebral Palsy Disabled

Maheshwari A. Ambewadikar<sup>1</sup>, Manasi R. Baheti<sup>2</sup>

Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, Aurangabad. Email: mahiambewadikar@gmail.com, manasibaheti@gmail.com

Speech has been the catalyst in the human development, being the natural mode to communicate, to interact. When it comes to the interaction between human and computer, it is accomplished via Automatic Speech Recognition (ASR). Many times, Speech is disordered, due to many reasons. The speech disorder in Cerebral Palsy disabled person motives speech impairment and fails to communicate. This paper presents the work for the speech recognition of Marathi digits and words of Cerebral Palsy (CP) Disabled person. The design and improvement (development) of the Automatic Speech Recognition (ASR) system in Marathi language, that will recognize the digits and some selected words used in daily communication, is the main task handled here.[2] The proposed system uses some techniques of the speech recognition, such as ZCR for endpoint detection, MFCC for the feature extraction of the pronounced word/digit, and DTW for the pattern matching. The database of the utterances of Cerebral Palsy (CP) persons is created and Automatic Speech recognition technique is applied for recognition. Such system can bridge the gap between Cerebral Palsy Disabled person and society as the system can be embedded in many applications like phone dialer, hands-free operations etc....

**Keywords :** Cerebral Palsy (CP), Automatic Speech Recognition (ASR), ZCR, MFCC, DTW.

## I. Introduction

Speech Recognition, also called as Automatic Speech Recognition (ASR) or

Computer Speech Recognition, implies understanding speech by the computer for performing any required task [1]. With the rapid development of the communication technologies, a reliable speech enabled technique is the necessity of the time [1].

A person can suffer from any form of the disability. Speech disability can come because of many reasons; this work considers the speech disorder due to Cerebral Palsy. Speech Recognition system for Cerebral palsy is this System that compares the utterance of the cerebral palsy person as input ,with reference utterance stored in database [2]. It is studied as of now that, currently no such system is available in the Marathi language. This paper presents the work that will develop the interface in Marathi language, to aid the CP disabled persons, to interact with the society, so that at least few, daily transactions in communication they can do with the help of the technology. Speech is variable in nature.[12] It is the ambitious task to recognize the speech. Basically it means talking to the computer/ or automated system, and recognize what is being said. It is very hard to recognize (identify) the speech of person with the disorder, like CP. Cerebral palsy is a disability, Where person having a Speech problem to produce Speech appropriately (properly) .Therefore such

persons have to rehash their words for communication.[8] It is aimed to develop an interface that will take input speech of speech disordered person and get it correctly recognized. It is a system where the stored speech references in the data base can be perceived (recognized) according to speech of CP disabled.[2].The system development includes creation of the database, in which digits in Marathi language and some selected words that are spoken in daily communication.[2]. Additional, algorithm are, Zero Crossing Rate (ZCR) utilized for the endpoint detection, Mel-Frequency Cepstral Coefficients (MFCC) using for the feature extraction, and Dynamic Time Warping (DTW) utilized for the pattern matching [2].

### Objective:

The Requirement for quick and broad application of Speech enabled systems is turning out to be evident Marathi is one of the native (regional) Indian languages. There are 12 vowels and 36 consonants present in Marathi languages. The objective of the research is to create the speech recognition system for Cerebral Palsy person to recognize digits in Marathi language.

### Motivation:

Cerebral palsy can affect a person's Speech ability to finely coordinate the muscles around the mouth and tongue

that are required (need) for speech. Few people with cerebral palsy will most likely be unable to produce any sounds, others might have the option to produce sounds but have difficulty controlling their movement enough to produce speech that is clear and comprehended by others.

There is no any Speech recognition system Developed in Marathi language for cerebral palsy people. A system existing in Marathi language so to overcome this language barrier want to develop this system.

To utilize of technology for disable people to bridge the gap between society & disable person.

## Literature Review:

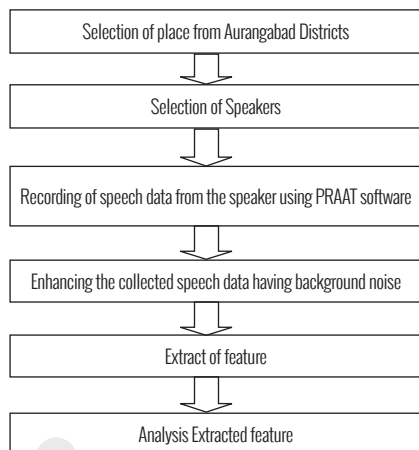
The Objective of this survey is to report the right research work and development done in speech recognition and their applications to HCI .The special emphasis is given to studies done with respect to application in Indian language.

In this the design of speech recognition system, relies upon the different issues: such as definition of various type of speech classes, speech representation, feature extraction techniques, speech classifiers, database, language models and performance evolution. The problems that exist in Automatic Speech Recognition (ASR) and the different techniques developed by various research workers. while literature review observe mostly used like Mel-frequency Cepstral Coefficient (Mfcc) for feature extraction using for Pattern Matching, Dynamic Time Warping (DTW) to solve these problems using various technique /Method and algorithms.

## II. Methodology and Experimental Work

### Methodology

Speech recognition system performs two principal activities: signal modeling and pattern matching .signal modeling represents process of changing (converting) speech signal into a set of parameters. Speech when recorded and saved to the system, becomes digitized. Speech wave part (components) are called as features but before the feature extraction we have a need of data collection speech samples from speaker and then extracted feature for further analysis. The technique followed by us for the proposed work is shown in figure.



**Fig. Methodology adopted for the proposed work**

## Implementation

### 1. Data collection Procedure

#### 1.1 Speaker Selection:

The speech data will be created from the essence (Native) speakers of Marathi language. The favored speakers will be from region of Aurangabad district. they would be comfortable with reading & speaking the Marathi language the speaker are classified on the basic gender.

#### 1.2 Speech Data Choice

I needed to have a restricted vocabulary to ensure the capture process was light weight, yet at the same time have enough verity for models trained on the data to potentially be utilize full for some applications I also needed the dataset to be usable in comparable ways to common proprietary collections like Digits this drove me to stop 10 common digits and Words as the core of our vocabulary.

#### 1.3 Speech Data Collection:

##### Data Collection

To access our strategy,a list of 10 digit short phrases in Marathi was compiled consisting of numbers & Words , commands to a typical information–access applications each entry was selected because it is either a single digit or a short phrase given our goal of high accuracy, small vocabulary speech recognition the vocabulary size was kept to a maximum limit of ten digit one target language were chosen <sup>[13]</sup>

#### 1.4 Recording Environment:

Speech Data Recovery is first step to words working of Speech Recognition system. The proficiency of recognition develop on the Speech Data used to prepare the system. At this moment the measures taken for gathering Speech information, to developed a ground-breaking Speech Recognition System as a grammatical feature. To achieve a high sound viewpoint, the account was done in the normal (ordinary) room without uproarious sound and impact of reverberation. The look at recurrence for all Recording was set to be 16000 Hz at the room temperature. The speakers were ask to sit in front from the amplifier with the separation of around 12-15 cm. The Speech data was assemble with the help of receiver ‘praat’ programming utilizing the single channel. The preprocessing was finished with the assistance of computerized speech laboratory facility (CSL). The computerized speech laboratory (CSL) is one of the examination framework for discourse and voice. Computerized speech laboratory (CSL) is one of the Input/output recording gadget for a PC, which has exceptional element for positive acoustic estimation.

## III. Experimental Work

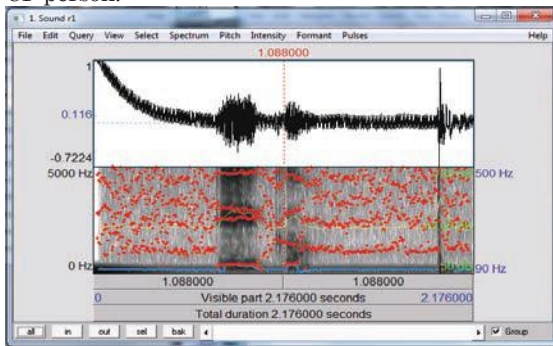
### III. I Creating Database-

Creating database is the most significant task. As speech is variant in nature, task oriented database is very basic

requirement in order to develop the system. Database was created for the proposed system.<sup>[5]</sup> Speech data acquisition is the initial step to words building of speech recognition system. The accuracy of recognition relies upon the speech data used to train the system.<sup>[6]</sup>

The data base is creates consist of selected digit in Marathi language used in daily life .digits from १ to १० (१,२,३,४,५,६,७,८,९,१०.) And word selected are Here used the cerebral palsy person's database which has scattered by age wise. The basic need of this database is to develop an automatic speech recognition system. The accuracy of speech recognition depends upon the nature of the discourse information gathered and the nature of training set data. The database is also created for this study. It is aim to developed speech recognition system that will recognize Marathi Digits.

Database of audio recordings of the persons with Cerebral Palsy and persons without cerebral palsy were taken <sup>[13]</sup>. First of all, Marathi digits 0 to 9 are recorded in the closed environment in the lab, with the help of the software Praat and by using the headphone with mike. This is very difficult task, as to have recordings at required frequency, at required amplitude, pitch and duration. Various utterances were taken before making it final for the recordings of the CP person.



**Fig. 1. Speech recorded in Praat for the word**

The proposed ASR system will recognize Marathi digits and words, spoken by Cerebral Palsy disabled person that is the system would be speaker independent .So, the ASR system needs to be trained with speech of lot of people so that it can handle variation in accent and speaking style. Recordings of five CP persons of different age group, both male and female were taken, each digit and word were uttered 10 times. Similarly recordings of same digits and words, with five persons with no disorder in speech were taken. In this recording, all speakers were local (native) speakers of Marathi language <sup>[14]</sup>.

In this recording, all speakers were local (native) speakers of Marathi language <sup>[14]</sup>. This group of speaker of is a combination of different gender, education level, and age. Secondly, the speakers with CP disability also belong to Marathi language as their mother tongue. Following are the words and digits in Marathi taken in to the work. This Database having two type of people one is normal and another

one Cerebral Palsy which is having Speech dieses but both are native person which is familiar with Marathi Language and both Speaker having same Digit to speak.

**Table 1. Marathi Digits.**

Sr. No.	Digit	Gender	Utterances
1	१	M and F	10
2	२	M and F	10
3	३	M and F	10
4	४	M and F	10
5	५	M and F	10
6	६	M and F	10
7	७	M and F	10
8	८	M and F	10
9	९	M and F	10
10	१०	M and F	10

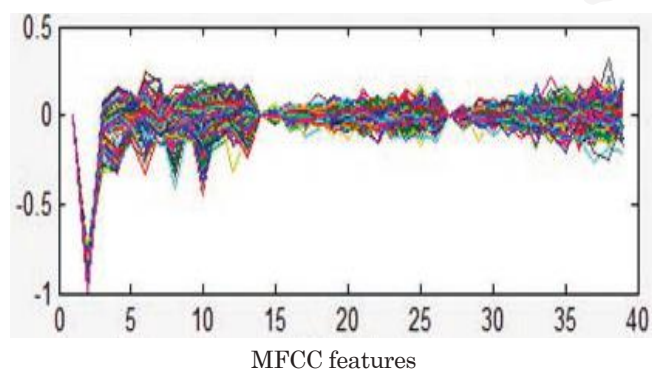
Automatic Speech Recognition can be achieved by a extensive variety of techniques. Basically, speech recognition starts with the digital sampling of speech. Most of the techniques include spectral analysis; e.g. LPC analysis (Linear Predictive Coding), MFCC (Mel Frequency Cepstral Coefficients), and many more. MFCC (Mel Frequency Cepstral Coefficient) is the next stage is recognition of phonemes, groups of phonemes and words <sup>[5]</sup>. After the MFCC (Mel Frequency Cepstral Coefficients) stage can be achieved by many processes such as DTW (Dynamic Time Warping). Automatic Speech Recognition system performs two fundamental operations: signal modeling and pattern matching <sup>[10]</sup>. Signal modeling represents process of converting speech signal into a set of parameters <sup>[9]</sup>. In this, features are extracted from the speech. Next step is Pattern matching is the task of finding parameter set from memory which closely matches the parameter set obtained from the input speech signal. The signal modeling involves with four basic operations: like spectral shaping, feature extraction, parametric transformation, and statistical modeling <sup>[9]</sup>. After database creation, next task was to apply feature extraction algorithm for looking into the details of the digitized speech signal. For this, most widely used technique of MFCC is applied and features are extracted.

### III. II Features Extraction

The first step in any automatic speech recognition system is to extract features that is Recognize (identify) the component of the audio signal [15] This will identify the syntactic (linguistic) content <sup>[7]</sup>. The major purpose of the MFCC process is to mimic the behavior of the human ears <sup>[8]</sup>. The method of filters the spaced linear at low frequencies and logarithmically at high frequencies have been used in order to capture the phonetically important characteristics of the voice <sup>[7]</sup>. MFCC extracts 13 (12parameters +1energy) features, these features can be applied with delta (26 features)

and double delta (39 features), for more recognition rate. [7] Following figure shows the window for the speech signal recorded using Praat. These audio recordings were called in Matlab to extract the MFCC features, following figure shows the plot of the extracted features [12].

MFCC (Mel Frequency Cepstral Coefficients) as the feature extraction technique The different values initialized during the computation of MFCC ((Mel Frequency Cepstral Coefficients) following.



After Feature Extraction Sampling frequency using 16000 and Hamming window is used with the window length of 25 millisecond per window given step time 10 millisecond and the minimum frequency is 0 (Zero) (lowest band edge mel filters(HZ) and maximum frequency 4000(the highest band edge of mel filters(HZ)set).

### III. IV Fast Fourier Transformation (FFT):

Fast Fourier Transformation converted each frame  $N$  samples from time domain in to the frequency domain (DFT), which is defined on the set of  $N$  samples  $\{X_n\}$ , as follows.

$$X_k = \sum_{n=0}^{N-1} X_n e^{-j2\pi kn/N}, \quad K=0,1,2\dots N-1$$

In general  $X_k$  are complex numbers and we only consider their absolute values (frequency magnitude). The resulting sequence  $\{X_k\}$  is interpreted as follow positive frequencies  $0 \leq f < F_s/2$ , correspond to values  $0 \leq n < N/2-1$ , while negative frequencies  $-F_s/2 < f < 0$  corresponds to as spectrum or periodogram. [15]

### III. V. Pattern Matching

After features are extracted, these needs to be matched. There are various pattern matching techniques, here, Dynamic Time warping (DTW) is a well know dynamic programming technique for finding the best alignment between two times series pattern is used [9]. DTW calculates the minimum distance between two words [12].

A local distance matrix is form for all the segment in the sample word and template word .This warping between two time series can then be used to find matching region between the two time series or to determine the similarity between the two time series [6]. Creation of the database, feature extraction and pattern matching is the main process in this

speech recognition [9].

As speech disorder is to be recognizing, it is very challenging task. The speech of the CP person is matched with the speech of the person without speech disorder. Further, the actual recognition rate is to be found out. But, as MFCC features are taken double delta time, the recognition rate will be increased. This type of system provides a prototype of an interface that will recognize the speech of speech disordered person with CP.

The significance of this work will be that, this can serve as the input for further linguistic study, also this can be the foundation to develop the various interfaces in Marathi language, such as phone dialer system, hands free operation system etc.

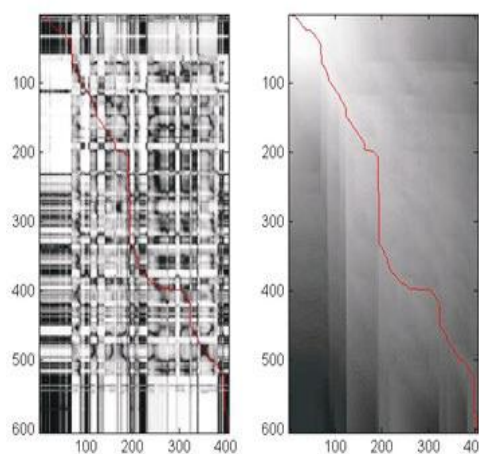


Fig. 3. DTW of two patterns of the same word.

### Future Scope

Once Digit Recognition is done we can develop different (various) applications for Cerebral Palsy (CP) people. Like digit recognition for mobile dialing, playing simple information. Using various combination of feature extraction techniques we can develop our database more accurately and efficient.

### Conclusion

The objective of this work is mainly to develop ASR System In Marathi Language for speech disorder of the CP person. In order to meet this objective, the first necessary step was to develop the database required. This system is intended at improving on the current Human-computer interface by developing a voice interface, which will prove to have so many advantages to the applications for the speech disorder. Today, even though much work in done regarding HCI in Speech Recognition; it is limited up to English language only. The concept of making these advanced systems reach up to speech disabled person and solve the problem of communication gap is the main problem handled over here.

## References

- 1] Ingyin Khine, Chaw Su “speech recognition system using MFCC and DTW” International Journal of Electrical, Electronics and Data Communication, ISSN(p): 2320-2084, ISSN(e): 2321-2950 Volume-6, Issue-7, Jul.-2018, <http://iraj.in>.
- 2] Mohd Hafidz Mohamad Jamil, S.A.R. Al-Haddad, CheeKyun, “A Flexible Speech Recognition System for Cerebral Palsy Disabled”, [https://link.springer.com/ chapter/10.1007/978-3-642-25327-0\\_5](https://link.springer.com/ chapter/10.1007/978-3-642-25327-0_5).
- 3] Mr. Kashyap Patel Dr. R. K. Prasad “Speech Recognition and Verification Using MFCC & VQ” International Journal of Advanced Research in Computer Science and Software Engineering [www.ijarcse.com](http://www.ijarcse.com) Volume 3, Issue 5, May 2013 ISSN: 2277 128X.
- 4] Shweta Bansal, “Isolated Word Speech Recognition System for Object identification”, International journal of Scientific and engineering research , Volume 4, ISSUE 1, January-2013.
- 5] Rahana fathema, Raseema P.E. “Gammatone-Cepstral-Coefficient-for Speaker”, International journal of scientific and engineering research, volume 4, Issue 10, October-2013 ISSN 2229-5518
- 6] Lindasalwa muda, mumtaj begam and i. elamvazuthi ”Voice Recognition Algorithms Using Mel-Frequency Cepsrtal Coefficient”, Journal of computer, value 2, Issue 3, March 2010, ISSN 2151-9617 [https://sites google.com/site/Journal of computing](https://sites.google.com/site/Journalofcomputing).
- 7] Prasanth P.S. “Speaker Recognition Using Vocal Tract Features”, International Journal of Engineering Inventions e-ISSN:2278 7461, P-ISSN: 2319-6491 Volume 3, Issue 1 (August 2013) PP. 26-30.
- 8] Risha Mal, R. K. Sharma ”Intoxicated speech Detection using Mfcc feature Extraction and Vector Quantization”, International Journal of electronic and electrical engineering ISSN 0974-2174. Volume 7. Number 3(2014). PP. 269-280.
- 9] Manish P. Kesarkar”Feature extraction for speech Recognition”, M.tech credit seminar Report, electronic systems group, EE Dept, IIT Bombay , Submitted November 2003.
- 10] S.A. Mohd yusof, P.M.Raj ”Speech Recognition Application Based on Malaysian Spoken Vowels Using Autoregressive Model of the vocal tract”, Processing of the international conference on electrical engineering and informatics institute technology Bondug, Indonesia. June 17-19, 2007.
- 11] Irfan Ahmad, Nasir Ahmd et.al “The Development of isolated words pashto Automatic Speech Recognition System”, <http://www.researchgate.net/publication/235726972> September 2012.
- 12] <https://www.abidibo.net/blog/2013/10/02/how-create-your-own-speech-recognition-application-tasker>.
- 13] pete warden “ Speech commands: A dataset for limited-vocabulary Speech Recognition”, google brain, mountain view, colifornia [petewarden@google.com](mailto:petewarden@google.com) April 2018, 1804,03209
- 14] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare, “Indian Language Speech Database: A Review”, Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, India, International Journal of Computer Applications (0975 – 888) Volume 47– No.5, June 2012.
- 15] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, “Design and devlopment of Spoken Marathi Isolated words Database for Agriculture Purpose and its Analysis” Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, April 2013.
- 16] Manasi Ram Bheti, Dr .S. C. Mehrotra “Speech Recognition System For Rural Applications Using Interactive Voice Response System”, Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, Feb-2016.

## About the Authors



**Maheshwari A. Ambewadikar** Working as M.Phil. Research Student Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra. She can be reached at [mahiambewadikar@gmail.com](mailto:mahiambewadikar@gmail.com)



**Dr. Manasi Ram Baheti** [Working as Assistant Professor Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra. She can be reached at [manasibaheti@gmail.com](mailto:manasibaheti@gmail.com)