International Conference on Advances in Materials, Computing and Communication Technologies
In Association with International Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Gujrati Speech Recognition Using Cellular Automata Algorithm

Vidya Gangadhar Dahake[1], Dr. Jagdish D. Kene[2]
[1]Department of ECE, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India
[2]Department of ECE, RCOEM, Nagpur, Maharashtra, India

ABSTRACT

The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. Speech Recognition can be defined as the process of converting speech signal to a sequence of words by means Algorithm implemented as a computer program. Speech processing is one of the exciting areas of signal processing. The goal of speech recognition area is to develop technique and system to develop for speech input to machine based on major advanced in statically modeling of speech, automatic speech recognition today find widespread application in task that require human machine interface such as automatic call processing. Communication among the human being is dominated by spoken language, therefore it is natural for people to expect speech interfaces with computer which can speak and recognize speech in native language. Machine recognition of speech involves generating a sequence of words best matches the given speech signal. The cellular automata algorithm is used for speech enhancement and reduction in noise. The researcher has studied the different Gujarati character from their pronunciation point. The main objective of the research is to dictate the Gujarati character pronounced by the user.

Index Terms- Speech enhancement, CA algorithm, Gujarati character, Feature Extraction, LPC, MFCC processor, pitch detection, Gujarati character

I. INTRODUCTION

Automatic Speech Recognition (ASR) also known as computer speech recognition is a process in which speech signal is converted into a sequence of words, other linguistic units by making use of an algorithm which is implemented as a computer program. The major objective with which ASR works is the development of the techniques and a system that enables then computers to recognize speech as input. In a speech recognition system we convert speech into text in which the text is the output of the speech recognition system which is equivalent to the recognized speech. Speech recognition applications have evolved over the past few years. These applications include voice search, call routing command and control, appliance control by voice, voice dialling, computer aided language learning, robotics and many more. Improved Linear Predictive Coding (LPC) coefficients of the frame are employed

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

in the feature extraction method. In the proposed speech recognition system, the static LPC coefficients + dynamic LPC coefficients of the frame were employed as a basic feature. After so many detailed study and optimization of ASR and various techniques of features extraction, accuracy of the system is still a big challenge. The selection of feature extraction techniques is completely based on the area of study. The cellular automata algorithm is used for speech enhancement and reduction in noise.

II. SPEECH ENHANCEMENT USING CA ALGORITHM

Cellular Automata (CA) were originally proposed by John von Neumann as formal models of self-reproducing organisms. CA generally used because of simplicity and parallel processing. CA has ability to perform complex operation and model behaviour of complex system in nature. It can be used in number of applications like mathematics, natural science, microstructure, computer science, biology, etc. CA has different forms like dynamic automata, tessellation automata, iterative automata, 1-dimensional automata, 2-dimensional automata, linear cellular automata, additive cellular automata. In proposed work, combination of iterative array and dynamic Cellular Automata is being implemented because in one iteration the weights are being updated and dynamically changes with respect to time. In this models each elementary device, a cell or site, can take some values and is updated at intervals according to a rule which expresses the actual value from its preceding value and that of its neighbours. Speech enhancement using CA algorithm has two phase first is training mode in which speech signal without noise is given input to the system and second phase is testing mode in which speech signal corrupted by adaptive background noise is applied to second part of system as shown in fig. 1. Here system adds some random noise to original signal, this signal along with original signal applied to CA block.

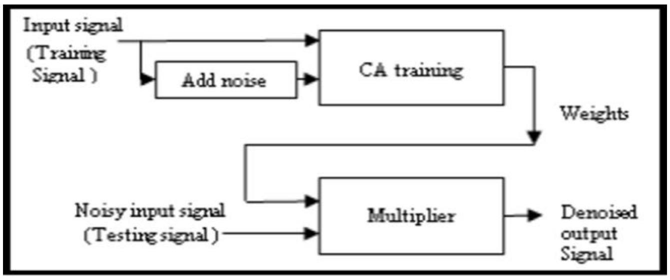
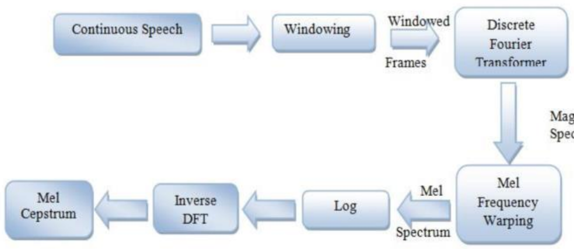


Fig.1 Block diagram of Speech enhancement using CA algorithm

III. METHODOLOGY

Linear predictive coding (LPC) method was introduced 60 years ago by Bishnu S. Atal and being used for speech vocal tracing because it represents vocal tract parameters and the data size are very suitable for speech compression. LPC analyses the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. In LPC system, each sample of the signal is expressed as a linear combination of the previous samples. This equation is called a linear predictor and hence it is called as linear predictive coding. But in LPC frequencies are weighted equally on a linear scale while the frequency sensitivity of the human ear is close to the logarithmic.

Feature Extraction Technique: Feature extraction is the main part of the speech recognition system. It is considered as the heart of the system. The work of this is to extract those features from the input speech (signal) that help the system in identifying the speaker. Feature extraction compresses the magnitude of the input signal (vector) without causing any harm to the power.



The above figure is the feature extraction diagram. In this, from one side we input the continuous speech signals for the process of windowing. In the process of windowing the disruptions which are present at the start as well as at the end of the frame are minimized. After this process, the continuous speech signal is converted into windowed frames. These windowed frames are passed into the discrete Fourier transformer which converts the windowed frames into magnitude spectrum. Now in the next step, spectral analysis is done with a fixed resolution along a subjective frequency scale that is the Mel-frequency scale which produces a Mel-spectrum. This spectrum is then passed to Log and then to inverse of discrete Fourier transform which produces the final result as Mel-Cepstrum. The Mel-Cepstrum consists of the features that are required for speaker identification. A few feature extraction techniques include:

Linear Predictive coding: LPC is a tool which is used for speech processing. LPC is based on an assumption: In a series of speech samples, we can make a prediction of the nth sample which can be represented by summing up the target signal's previous samples (k). The production of an inverse filter should be done so that is corresponds to the formant regions of the speech samples. Thus the application of these filters into the samples is the LPC process.

IV. DEVELOPMENT OF TRUE TYPE FONTS (TTF)

The developed "mansi.ttf" is a True Type Font in Gujarati, so the installation process does not require ATM. Copying the mansi.ttf only into font directory of the windows will complete the installation process.

In Gujarati other .ttf fonts are available, but one has to purchase it for its regular use.

I have also studied some free Gujarati true type fonts which are available on internet, but some of it are not following the standard A S D format.

I have also identified some true type fonts which follows A S D format but they are not supporting some special characters i.e. S Þ ß ö4 ô ö ÷

Due to all above-mentioned reason, the researcher was guided to develop such a free True Type Font named "mansi" which satisfy all the requirements of the researcher. To develop the "mansi.ttf" the researcher has used the Font Creator Program– 4.0 of High Logic.

Algorithm to develop font

During the development process of the Gujarati characters researcher is able to identify the steps to construct the font in any language. The researcher is presenting these steps in the form of an algorithm as follow:

(1) Identify the character set for the specific language for which we are developing the font.
(2) Prepare the sample layout for each character for the language.
(3) Scan each character and store it into a .bmp file separately.
(4) Using some picture editor software (i.e. Photoshop or CorelDraw) enhance the curve and clarity of the scanned image for each character.
(5) Determine the alphabet layout for each character and assign specific ASCII to each character.
(6) Select any font creator program.
(7) Pick up any character and map it to specific glyphs.
(8) Align the character at the pre-determined base line and resize the font to a predetermined size i.e. height.
(9) Set the space after character in the glyphs.
(10) Repeat the above steps (7) to (9) for each character of the language.

Character list of developed Gujarati font named "mansi.ttf"

The letters of alphabet are divided into two categories. The letters, which can be pronounced independently by them, are called vowels. The letters, which cannot be pronounced independently by them, are called consonants. The research has divided the Gujarati character set into four different categories as 34 consonants, 11 vowels, 10 digits and 106 special characters.

List of Consonants

S B U 3 R K H h 8 9 0 - 6 T Y N W G 5 O A E D I Z , J ; X Ø C / Ù 7

List of Digits

_ ! Z # $ 5 & * ( )

Considering all these points in mind the researchers has tryout the all these measures and developed the fonts. These measurements are summarized as follow:

Name	Measurements
Base Line	0
Mean Line	1350
Cap Line	1490
Ascender Line	1950
Descender Line	-530
Leading	50

Using all the above-mentioned measurement the researcher has developed the fonts using Font Creator programme

V. DEVELOPMENT OF SPEECH EDITOR

In this approach, instead of having to type commands, users could choose the command form a list, called a menu. Menu gives the list of assorted command, from which user is able to choose any command . Researcher has developed the prototype model for the interaction with computer i.e. character detection. For the character detection there must be one interface, which takes the input character voice through microphone and sends the created wave file to the routine for the further processing. So in the model

researcher has kept the front end, which is second component of the entire model. This front end is designed and developed by the researcher named "ckksedit".

"ckksedit" is a word processing program which can be used for a lot many work involving creating and managing text. A lot many features are added in this program, which is developed using VC++ of Visual studio 6.0. "ckksedit" speech editor offers a range of features usually found in PC based word processors and provides speech feedback in character mode. The editor software provides a means of editing existing rich text format (RTF) files or creating new ones. The editor runs under windows environment.

"ckksedit" is a purpose designed speech editor which gives the facility to format the text, play .wav file, record a .wav file and provides specific utilities suitable for speech processing i.e. detection of a character from Gujarati alphabet. After taking the input in the detection mode the character wave file is created named "ckkChar.wav" and it sends its speech output to Mat lab routine named "ckkchar". This routine compare the characteristics of the spoken character with the master database and return the highly probably matching character back to the editor, editor accept the character and represents the corresponding character on the screen in the textual form.

VI. CREATION OF SPEECH FILE

Researcher has set the following parameter and then recording of the speech file is performed for all the four users. Gujarati language consists of 12 vowels, 34 consonant, 10 digits and many more special characters. For the experimental work researcher has selected only consonant. While study the characteristics of the Gujarati character, the researcher has identified the different five characters for the experimental work. S4 R4 Z4 T4 D.

File type	.wav
Sampling frequency	11025Hz
No. of channel	Stereo(1ch)
Quantization bits	16 bits

Table-4.9 Recording parameter

To get more detail the researcher feels to study more samples for these five Gujarati characters i.e. (S4 R4 Z4 T4 D.) And hence again he has pronounced the each character five times from the same speaker and prepared the different speech file for each speaker and stored these entire files in their respective directories. The file name are given as ka101, ka102, ka103, ka104, ka105, ch101, ch102, ch103, ch104, ch105, ra101, ra102, ra103, ra104, ra105, t101, t102,t103, t104, t105, ma101, ma102, ma103, ma104 and ma105 for speaker 1 i.e. (CKK). The same pattern is repeated for the second speaker and the different speech files are created. The name of the file follows the same pattern except that 101, 102, 103, 104, 105 becomes 201, 202, 203, 204 and 205 respectively. For the third speaker it is 301, 302 and so on. The process is repeated for fourth, fifth and sixth speaker. The speech file is available in the same CD-ROM in the folder “master data”.

VII. PITCH DETERMINATION

Pitch is defined as that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale. One of the advantages of using pitch analysis or pitch detection is its development of a robust and accurate pitch detector has been the focus of much research in the area of speech analysis. End Point Detection algorithm must take a number of special situations into account such as

- Words, which begin or end with low-energy phonemes (weak fricatives)
- Words, which end with an unvoiced plosive.
- Word, which end with a nasal.
- Speakers ending words with a trailing off in intensity or a short breath (noise)

Cepstrum is the IDFT of the logarithm of the magnitude of the DFT of the signal. The cepstral plot gives a peak at the pitch period for voiced speech and there is no substantial peak in the cepstrum plot for unvoiced speech.

The little limitation of Cepstrum-Liftering method is that it is subjected to noisy disturbance. This is because, for noisy speech cepstrum, the peak produced from the source is dominated by that produced from noise. Therefore, the liftering cut-off point must be estimated appropriately according to speaker gender. The following figure represents the cepstrum analysis algorithm for pitch detection.

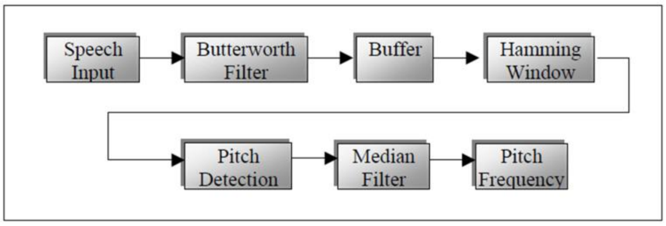


Figure – Cepstrum analysis steps for pitch detection

In final step, we convert the log Mel spectrum back to time. The result is called the Mel-Frequency Cepstral Coefficients (MFCC). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the Mel spectrum coefficients (and so their logarithm) are real numbers, we can convert them to the time domain using the Discrete Cosine Transform (DCT). Therefore we can calculate the MFCC's.

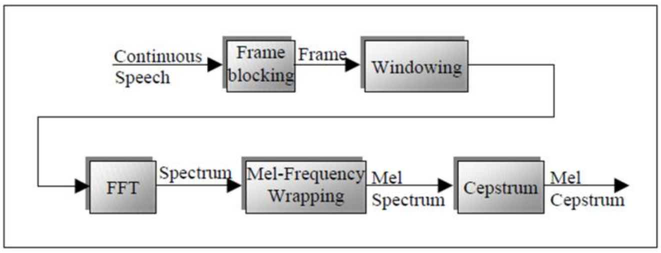


Figure – Block diagram of the MFCC processor

VIII. CONCLUSION

The researcher has studied the different Gujarati character from their pronunciation point and determines to concentrate only on five characters. To study the characteristics of the Gujarati character, the pronounced speech was required. So he has performed the recording from the different six speakers and stored the output in the form of wave file. For the creation of speech wave file he has selected three female speakers and three male speakers. The recording was done in the "ckksedit" editor. While creating the wave file, the researcher has taken care about the three base parameters i.e. (1) speaker characteristics (2) recording site and platform and (3) storing mechanism of speech file. All speech files are stored in "master data" in CD-ROM.

When the character is pronounced and recorded in the wave file, it consists some unnecessary voice, to remove such voice researcher has used "end point detection" algorithm, which remove all such unnecessary voice from the speech file, and generates another file.

The researcher has used total 125 wave files for various pronounced utterance of Gujarati character from different six speakers. He has applied the algorithm on all these files and stored the extracted file with appropriate file name. All these files are available in "files after end point detection" folder in the CD-ROM.

For gender detection, researcher has used pitch detection algorithm. For character detection, researcher has designed and develops an algorithm, applied this code on the stored 125 speech wave file.

By analysis researcher has generated the success rate of the algorithm for each character and the result is tabulated in the following table:

Speaker	CKK	ADA	PAA	HCK	KPM	Total out of 25	Success %
ક	5	3	1	1	5	15	60
ચ	4	1	3	1	5	14	56
ર	5	4	4	0	3	16	64
ત	1	0	0	0	0	01	04
મ	5	5	4	4	5	23	92

Table – 5.9 Percentage wise success rate of each character

IX. REFERENCES

[1]. Mahesh B. Adhav, Jagdish D. Kene" A Novel Approach of Speech Enhancement by Implementing Cellular Automata Algorithm" International Journal of Computer Applications (0975 – 8887) Volume 145 – No.8, July 2016

[2]. Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar" A Review on Speech Recognition Technique" International Journal of Computer Applications (0975 – 8887)Volume 10– No.3, November 2010"

[3]. Desai Vijayendra and Dr.Vishvjit K. Thakar, "Neural Network based Gujarati Speech Recognition for Dataset Collected by in-ear Microphone", in Proceedings of 6th International Conference on Advances in Computing & Communications ICACC 2016, Cochin, ISSN: 1877-0509, 6-8th September 2016, pp.668-675.

[4]. R. Rosenfeld, "Two decades of statistical language modeling: Where do we go from here?," IEEE Proceedings, vol. 88, no. 8, pp. 1270–1278, 2000.

[5]. A.Srinivasan, K.Srinivasa Rao, K.Kannan and D.Narasimhan, "Speech Recognition of the letter 'zha'(H) in Tamil Language using HMM", International Journal of Engineering Science and Technology, Vol.1(2),2009, pp.67-72

[6]. Geeta Nijhawan and Dr. M.K Soni, "Real Time Speaker Recognition System for Hindi Words", International Journal of Information Engineering and Electronic Business, Vol. 6, DOI: 10.5815/ijieeb.2014.02.04, April 2014, pp. 35-40.

[7]. Hazrat Ali, An Jianwei and Khalid Iqbal, "Automatic Speech Recognition of Urdu Digits with Optimal Classification Approach", International Journal of Computer Applications, Vol.118, No.9, May 2015, pp. 1-5.

AUTHORS

First Author – Vidya Gangadhar Dahake, Department of ECE , Sri Satya Sai University of Technology & Medical Sciences, Sehore, India, vidyadahake@gmail.com

Second Author – Dr. Jagdish D Kene, Department of ECE,RCOEM Nagpur.