ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2015

Design and Implementation of Silent Pause Stuttered Speech Recognition System

V.Naveen Kumar[1], Y Padma Sai[2], C Om Prakash[3]

Project Engineer, Dept. of ECE, VNRVJIET, Bachupally, Hyderabad, Telangana, India [1]

Professor & Head of the Dept, Dept. of ECE, VNRVJIET, Bachupally , Hyderabad, Telangana, India [2]

PG Student, Dept. of ECE, VNRVJIET, Bachupally , Hyderabad, Telangana, India [3]

ABSTRACT: Humans use speech as a verbal means to express their feelings, ideas, and thoughts for communication. In this world, there is 1% of the population having the problem of speech dysfluency**.** Stuttering is one such disorder in which the fluent flow of speech is disrupted by occurrences of dysfluencies such as silent-pauses, repetitions, prolongations, interjections and so on. Eliminating such dysfluencies would be helpful for the people with speech disorder to easily transfer their ideas and communicate easily. This paper proposes a system which will remove silent pauses from the speech and produce the corrected speech format which can be easily understood.

KEYWORDS: Stuttering, Silence Removal, MFCC, Dynamic Time Warping, Speech Recognition.

I.INTRODUCTION

Humans use various methods for communication purpose, in which speech is the vocalized form of human communication. Necessity to communicate with the machines led to the technique called speech recognition. Speech recognition is the process of identifying the spoken speech. Though speech recognition technology improved in recent decades more challenges are waiting for it, one such challenge is speech recognition for stuttered speech. Speech stuttering also known as dysphemia and stammering is a disorder that affects the fluency of speech [4]. It occurs in about 1% of the population and has found to affect four times as many males as females. Stuttering is one such disorder in which the fluent flow of speech is disrupted by occurrences of dysfluencies such as silent-pauses, repetitions, prolongations, interjections and so on [2].

Stuttering is the subject of interest to researchers from various domains like speech physiology, pathology, acoustic and signal analysis. In conventional stuttering systems major work is done on stuttering assessment process in which the recorded speech is transcribed and dysfluencies like silent-pauses, repetitions, prolongations etc are identified [3].
The objective of the work is to develop a system capable of finding the dysfluencies in stuttered speech and identify the corrected speech. This helps people with speech disorder to easily communicate and exchange their ideas.

II. STUTTERED SPEECH RECOGNITION SYSTEMS

Throughout the human history, speech has been the most dominant and convenient means of communication between people. Today, speech communication is not only for face-to-face interaction, but also between individuals at any moment, anywhere, via a wide variety of modern technological media, such as wired and wireless telephony, voice mail, satellite communications and the Internet. The recognition accuracy of a machine is, in most cases, far from that of a human listener, and its performance would degrade dramatically with small modification of speech signals or speaking environment. Due to the large variation of speech signals, speech recognition inevitably requires complex algorithms to represent this variability. A typical speech signal consists of two main parts: one carries the speech information, and the other includes silent or noise sections that are between the utterances, without any verbal information [8]. The verbal part of speech can be further divided into two categories as voiced speech and unvoiced speech. Being able to distinguish between the two is very important for stuttered speech recognition. The first speaker's characteristics have to be changed gradually to those of the second speaker; therefore, the pitch, the duration, and the spectral parameters have to be extracted from both speakers.

Then natural-sounding synthetic intermediates have to be produced. It should be emphasized that the two original signals may be of different durations, may have different energy profiles, and will likely differ in terms of many other vocal characteristics. Unvoiced speech sections are generated by forcing air through a constriction formed at a point in the vocal tract (usually toward the mouth end), thus producing turbulence. The characteristic features for voiced and unvoiced speech determination are zero crossing rate and energy. Energy is used for removing silent pause stuttering which is considered as the unvoiced speech from the speech. The stuttered speech recognition is mainly carried out in two phases namely training and testing. The major phases of classification system are pre-emphasis, stutter removal, segmentation, feature extraction, VQ codebook generation and score matching.

III.EXPERIMENT

Analysis of stuttered speech and recognition is done as described below.

1 Pre-Emphasis:

In general speech waveform suffers from additive noise. The performance of automatic speech recognition systems degrade greatly when speech is corrupted by noise. In order to enhance the accuracy and efficiency of the extraction process, speech signals are pre-processed [5]. Pre-emphasis is performed by filtering the speech signal with first order FIR filter, which takes the following form:

$$H(z) = 1-k*z^{-1} \qquad (0.9<k<1)$$

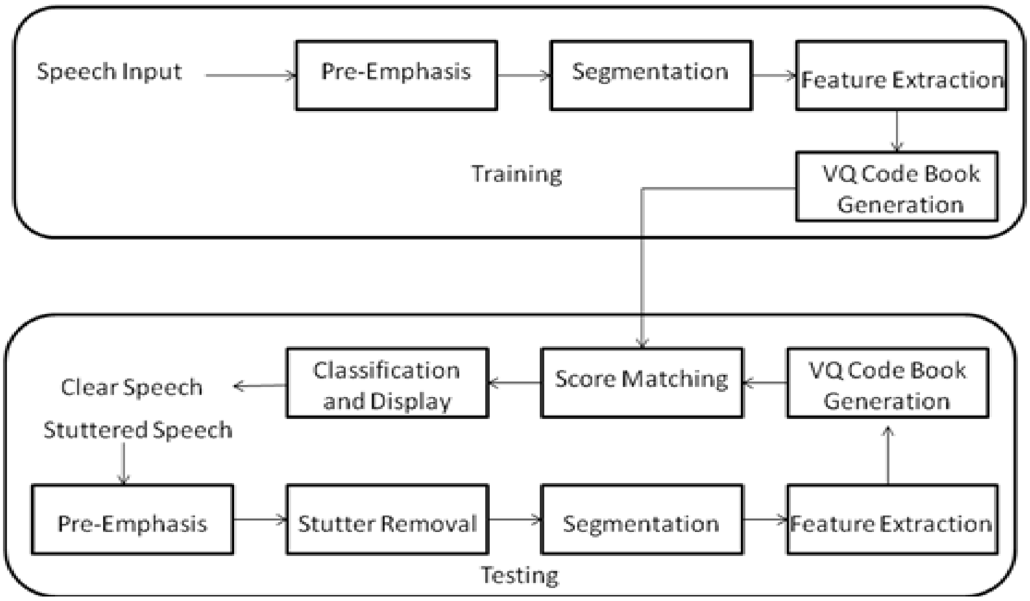


Fig.1: Block diagram of the system

2 Stutter Removal:

Stuttering is a speech disorder with many definitions characterized by certain types of speech dysfluencies. The different dysfluency classes are: broken words; sound prolongations; word repetitions; syllable repetitions; interjections; and phrase repetitions [3]. This paper proposes the use of speech recognition technology to identify the silent paused stuttered speech.

A verbal speech signal can be categorized into two as voiced speech and unvoiced speech. Being able to distinguish between voiced and unvoiced speech is very important for speech signal analysis, which can be determined by characteristic features like energy and zero crossing rate. Energy feature of the speech signal is employed for determining voiced and unvoiced speech.

The energy of the unvoiced speech is less than the voiced speech. The energy of speech sample which is below ten percent of the maximum energy to be considered as unvoiced speech and it is removed. Therefore, stuttered speech is transformed to a stutter free speech signal.

3 Framing:

Analyzing a stationary signal is simple and easy compared to continuously varying signal. The speech signal is continuously varying but from a short time point of view it is stationary, this is from the fact that glottal system cannot change immediately and research states that speech is typically stationary in the window of 20ms. Therefore the signal is divided into frames of 20ms which corresponds to n samples:

$$n=t_{st}f_s$$

In speech processing it is often advantageous to divide the signals into frames to achieve stationary.

4 Feature Extraction:

To identify a speech signal features should be matched with the previous signal or upcoming signal. Hence feature extraction is performed to convert speech signal to some types of parametric representations for further analysis. There are several feature extraction techniques namely Linear Predictive Coefficients (LPC), Linear Predictive Cepstral Coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction cepstra (PLP) [4].

MFCC is one of the successful feature extraction methods in speech dysfluency classification [1]. MFCC is used as it is based on the known variations of the human ear's critical bandwidths, with frequency filters spaced linearly at low frequencies and logarithmically at high frequencies to capture important characteristics of speech. This is expressed in the mel-frequency scale; which is a linear spacing below 1000Hz and a logarithmic spacing above 1000Hz. The approximate formula to compute the Mel's for a given frequency f in Hz is given by:

$$\text{Mel}(f) = 2595 * \log10\,(1+ (f/700)) \qquad (1)$$

The MFCC features are calculated using the following process:

4.1 Windowing:

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities by using the window to taper the signal to zero at the beginning and end of the frame. Windowing is a point wise multiplication between framed signal and the window function. A good window function has a narrow main lobe and low side lobe levels in their transfer functions. The hamming window is applied to minimize the spectral distortions and discontinuities. The hamming window coefficients are estimated as:

$W(n) = 0.54- 0.46 \cos (2\prod (n/N)) \; (0 \leq n \leq N) \qquad (2)$

4.2 Fast Fourier Transform (FFT):

The speech signal can be analysed much better in frequency domain. Thus, FFT is applied on the windowed signal which is essentially still a DFT for transforming discrete time domain signal into its frequency domain [5]. The difference is that FFT gives more efficient and faster computations which are given by the equation:

$Y(w) = FFT (h(t) * X(t)) = H(w) * X(w) \quad (3)$

4.3 Mel Frequency Wrapping:

One way to more concisely characterize the signal is through filter banking [6]. The frequency ranges of interest are divided into N bands and measure the overall intensity in each band. Intensity in each band is measured by simply adding up all the values in the range, or compute "power" measure by summing the squares of the values [4]. To agree better with the human perceptual capabilities mel-frequency scale is used which follows a linear spacing below 1000Hz and a logarithmic spacing above 1000Hz.

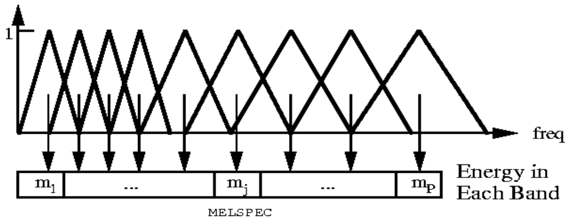


Fig.2: MEL scale filter bank

4.4 Discrete Cosine Transform (DCT):

The last process in Mel-Filter feature extraction is to apply inverse transform to obtain the enhanced speech signal. Since speech signal is not present in the entire transform coefficient and to obtain original signal DCT is applied. DCT provides higher energy compaction as compared to DFT [7]. Unlike DFT the DCT coefficients are real and there is no phase component. Hence DCT is a good choice for speech enhancement. With the values from each filter band given, cepstrum parameter in Mel scale can be estimated and MFCC features are obtained.

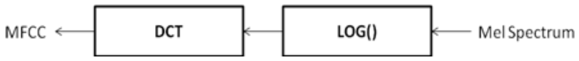


Fig. 3: Mel Cepstrum Coefficients

4.5 Vector Quantization:

Vector Quantization (VQ) is an efficient and simple approach for data compression. It is used to preserve the prominent characteristics of data [5]. VQ is one of the ideal methods to map huge amount of vector from a space to a predefined number of clusters, each of which is defined by its central vector or centroid. One of the key point of VQ is to generate a good codebook such that distortion between the original signal and the reconstructed signal is the minimum. Various techniques to generate codebook are available. The method most commonly used to generate codebook is the K-means algorithm [4]. The K-means algorithm is a straightforward iterative clustering algorithm that partitions a given dataset into user specified number of clusters K. In brief, the K-means algorithm is composed of the following steps:

1. Clusters the data into k groups where k is predefined.
2. Selects k points at random as cluster centers.
3. Assigns objects to their closest cluster center according to the Euclidean distance function.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeats steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

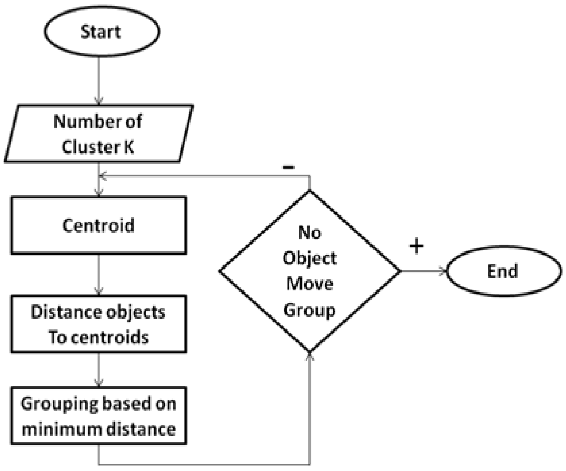

Start
Number of Cluster K
Centroid
Distance objects To centroids
Grouping based on minimum distance
No Object Move Group
-
+
End

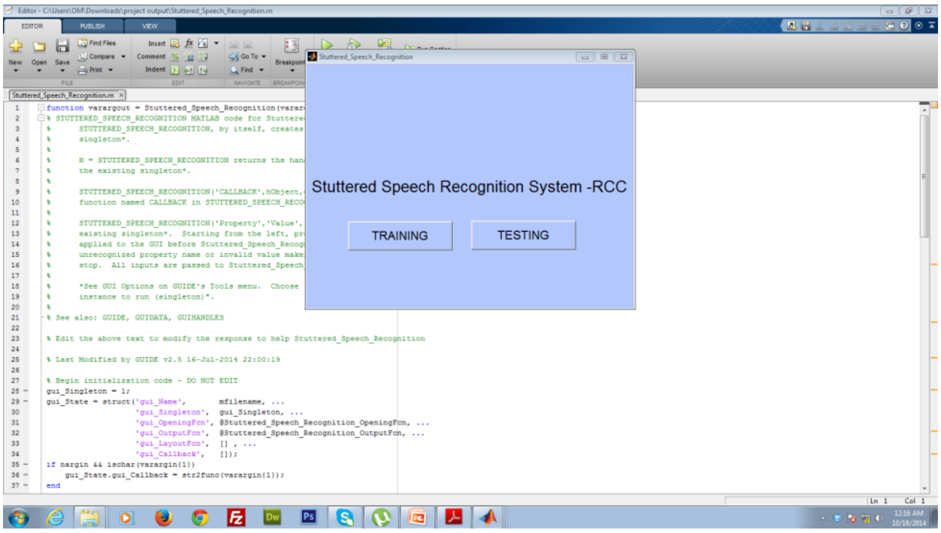
Fig. 4: Steps in K-means algorithm

4.6 DTW Score Matching:

Comparing the template with incoming speech might be achieved via a pair wise comparison of the feature vectors in each. The total distance between the sequences would be the sum or the mean of the individual distances between feature vectors. The problem with this approach is that if constant window spacing is used, the lengths of the input and stored sequences are unlikely to be the same. The Dynamic Time Warping algorithm achieves this goal; it finds an optimal match between two sequences of feature vectors which allows for stretched and compressed sections of the sequence [8]. In time series analysis, dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed [8].

IV. RESULT AND DISCUSSION

A speech recognition system capable of finding the dysfluencies in a silent paused stuttered speech and produce the corrected speech has been developed. MATLAB software is used for developing stuttered speech recognition and correction system. MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming environment. It has powerful built-in routines that enable a very wide variety of computations. It also has easy to use graphics commands that make the visualization of results immediately available. Specific applications are collected in packages referred to as toolbox. There are toolboxes for signal processing, symbolic computation, control theory, simulation, optimization, and several other fields of applied science and engineering.

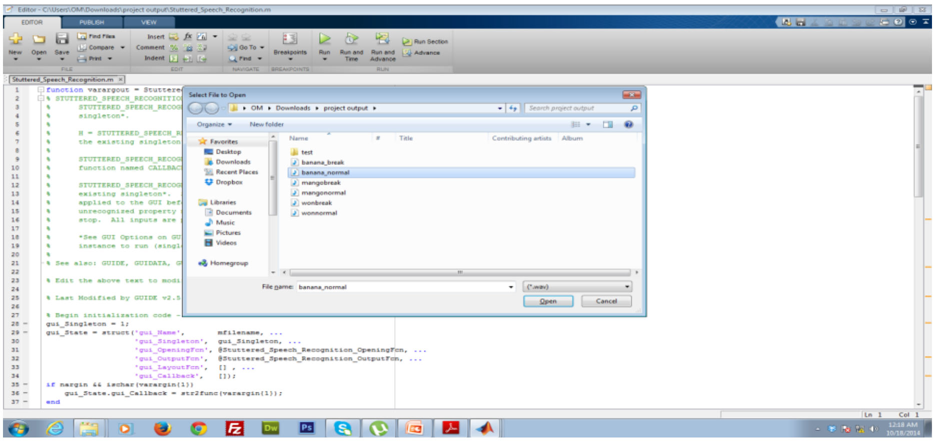

Editor - C:\Users\OM\Downloads\project output\Stuttered_Speech_Recognition.m
EDITOR
PUBLISH
VIEW
New
Open
Save
Find Files
Compare
Print
Insert
Comment
Indent
Go To
Find
Breakpoints
FILE
EDIT
NAVIGATE
BREAKPOINTS
Stuttered_Speech_Recognition.m
Stuttered_Speech_Recognition
Stuttered Speech Recognition System -RCC
TRAINING
TESTING
Ln 1 Col 1

```
function varargout = Stuttered_Speech_Recognition(vararg
% STUTTERED_SPEECH_RECOGNITION MATLAB code for Stuttere
%      STUTTERED_SPEECH_RECOGNITION, by itself, creates
%      singleton*.
%
%      H = STUTTERED_SPEECH_RECOGNITION returns the hand
%      the existing singleton*.
%
%      STUTTERED_SPEECH_RECOGNITION('CALLBACK',hObject,
%      function named CALLBACK in STUTTERED_SPEECH_RECO
%
%      STUTTERED_SPEECH_RECOGNITION('Property','Value',
%      existing singleton*.  Starting from the left, pr
%      applied to the GUI before Stuttered_Speech_Recog
%      unrecognized property name or invalid value make
%      stop.  All inputs are passed to Stuttered_Speech
%
%      *See GUI Options on GUIDE's Tools menu.  Choose
%      instance to run (singleton)*.
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help Stuttered_Speech_Recognition

% Last Modified by GUIDE v2.5 16-Jul-2014 22:00:19

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @Stuttered_Speech_Recognition_OpeningFcn, ...
                   'gui_OutputFcn',  @Stuttered_Speech_Recognition_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end
```

Fig. 5 MATLAB Simulation of the system

In the fig 1, it shows the MATLAB setup and the system GUI system for stuttered speech recognition system.

The acoustic model parameters of the speech units are estimated using training data. Language models are obtained from the collected large database with script files.


Editor - C:\Users\OM\Downloads\project output\Stuttered_Speech_Recognition.m

Stuttered_Speech_Recognition.m

```
function varargout = Stuttere
% STUTTERED_SPEECH_RECOGNITIO
%       STUTTERED_SPEECH_RECOG
%       singleton*.
%
%       H = STUTTERED_SPEECH_R
%       the existing singleton
%
%       STUTTERED_SPEECH_RECOG
%       function named CALLBAC
%
%       STUTTERED_SPEECH_RECOG
%       existing singleton*.
%       applied to the GUI bef
%       unrecognized property
%       stop.  All inputs are
%
%       *See GUI Options on GU
%       instance to run (singl
%
% See also: GUIDE, GUIDATA, G

% Edit the above text to modi

% Last Modified by GUIDE v2.5

% Begin initialization code -
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @Stuttered_Speech_Recognition_OpeningFcn, ...
                   'gui_OutputFcn',  @Stuttered_Speech_Recognition_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end
```

Select File to Open

OM ▸ Downloads ▸ project output ▸

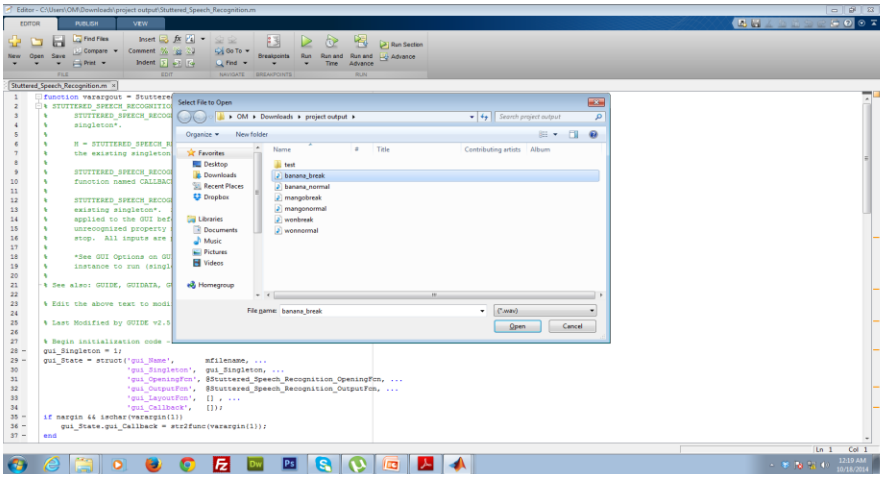
test
banana_break
banana_normal
mangobreak
mangonormal
wonbreak
wonnormal

File name: banana_normal

(*.wav)

Fig. 6 MATLAB Demonstration of Training Phase of the system

In the fig 2, training phase is depicted which comprises of speech data collection, feature extraction and also extraction and storage of the model parameters from the features of the training data.


Editor - C:\Users\OM\Downloads\project output\Stuttered_Speech_Recognition.m

Stuttered_Speech_Recognition.m

```
function varargout = Stuttere
% STUTTERED_SPEECH_RECOGNITIO
%       STUTTERED_SPEECH_RECOG
%       singleton*.
%
%       H = STUTTERED_SPEECH_R
%       the existing singleton
%
%       STUTTERED_SPEECH_RECOG
%       function named CALLBAC
%
%       STUTTERED_SPEECH_RECOG
%       existing singleton*.
%       applied to the GUI bef
%       unrecognized property
%       stop.  All inputs are
%
%       *See GUI Options on GU
%       instance to run (singl
%
% See also: GUIDE, GUIDATA, G

% Edit the above text to modi

% Last Modified by GUIDE v2.5

% Begin initialization code -
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @Stuttered_Speech_Recognition_OpeningFcn, ...
                   'gui_OutputFcn',  @Stuttered_Speech_Recognition_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end
```

Select File to Open

OM ▸ Downloads ▸ project output ▸

test
banana_break
banana_normal
mangobreak
mangonormal
wonbreak
wonnormal

File name: banana_break

(*.wav)

Fig .7 MATLAB Demonstration of Testing Phase of the system

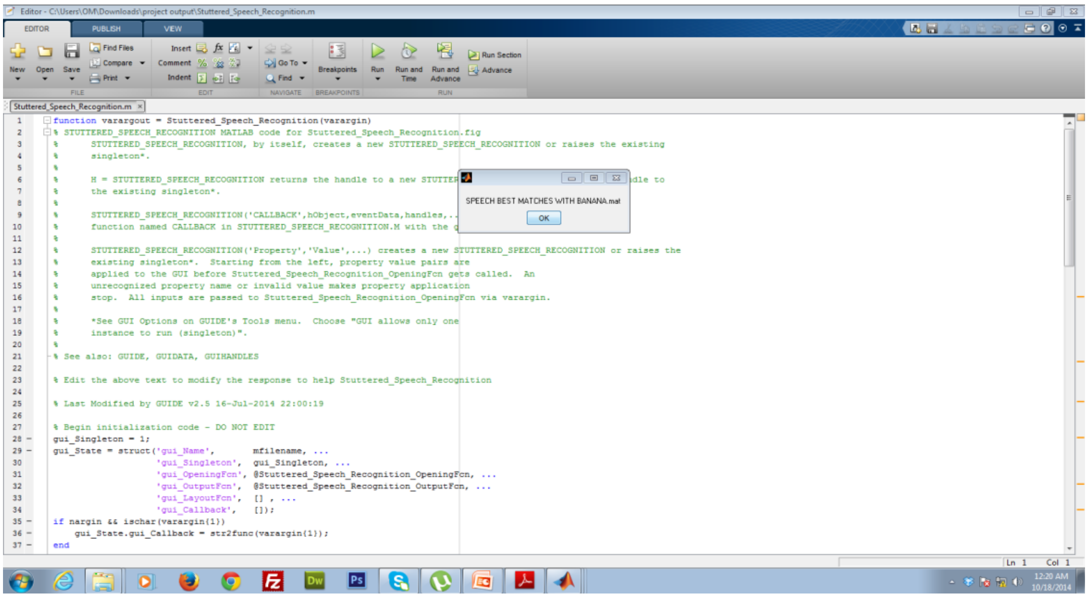
In Fig 3, In this testing process Stuttered speech is given to the system. The system eliminates the stuttering from the speech and extracts MFCC features and compares it with the training database of the fluent speech. After matching the stuttered speech with the fluent speech using Dynamic Time Warping. The identified speech is displayed, and then the identified sound is obtained from the speaker.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2015



```
function varargout = Stuttered_Speech_Recognition(varargin)
% STUTTERED_SPEECH_RECOGNITION MATLAB code for Stuttered_Speech_Recognition.fig
%      STUTTERED_SPEECH_RECOGNITION, by itself, creates a new STUTTERED_SPEECH_RECOGNITION or raises the existing
%      singleton*.
%
%      H = STUTTERED_SPEECH_RECOGNITION returns the handle to a new STUTTER                                   dle to
%      the existing singleton*.
%
%      STUTTERED_SPEECH_RECOGNITION('CALLBACK',hObject,eventData,handles,..
%      function named CALLBACK in STUTTERED_SPEECH_RECOGNITION.M with the g
%
%      STUTTERED_SPEECH_RECOGNITION('Property','Value',...) creates a new STUTTERED_SPEECH_RECOGNITION or raises the
%      existing singleton*.  Starting from the left, property value pairs are
%      applied to the GUI before Stuttered_Speech_Recognition_OpeningFcn gets called.  An
%      unrecognized property name or invalid value makes property application
%      stop.  All inputs are passed to Stuttered_Speech_Recognition_OpeningFcn via varargin.
%
%      *See GUI Options on GUIDE's Tools menu.  Choose "GUI allows only one
%      instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help Stuttered_Speech_Recognition

% Last Modified by GUIDE v2.5 16-Jul-2014 22:00:19

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @Stuttered_Speech_Recognition_OpeningFcn, ...
                   'gui_OutputFcn',  @Stuttered_Speech_Recognition_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end
```

SPEECH BEST MATCHES WITH BANANA.mat

OK

Fig .8 MATLAB identifying the corrected speech

VI.CONCLUSION

In this paper a new approach for correction and recognition of silent paused stuttered speech is presented. Stuttering is eliminated by considering the fact that voiced speech has more energy than the unvoiced speech. The feature extraction was performed using MFCC algorithm. The VQ code book is generated by clustering the training features vectors of the dysfluent speech and then stored in the database. In this method, the K-means algorithm is used for clustering purpose. DTW algorithm was used to match the dysfluent speech with the database. Finally the silent paused stuttered speech is corrected and stutter free speech is recognized.

Stuttered speech recognition and correction system is successfully developed. And this system is used to clearly understand the words uttered by a person with speech disorder. The current system is employed only for isolated silent pause stuttering word. The system can be further improved for complete sentences and also for multi modal stuttering.

REFERENCES

1. Chong Yen Fook, Hariharan Muthusamy, Lim Sin Chee, Sazali Bin Yaacob, Abdul Hamid bin Adom “Comparison of Speech parameterization techniques for the classification of speech disfluencies”, Turk J Elec Eng & Comp Sci (2013) 21: 1983 – 1994.
2. K.M. RaviKumar, R.Rajagopal, H.C.nagaraj, “An Approach for Objective Assessement of Stuttered Speech Using MFCC Features”, DSP Journal, Volume 9, Issue 1, June, 2009.
3. Lim Sin Chee, Ooi Chia Ai, Sazali Yaacob, “Overview of Automatic Stuttering Recognition Syste,” in Proceedings of the International Conference on Man-Machine System (ICoMMS) 11 - 13 Octomber 2009, Batu Ferringhi, Penang, Malaysia.
4. P.Mahesha and D.S. Vinod, “An Approach for Classification of Dysfluent and Fluent Speech using K-NN and SVM,” International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.6, December 2012.
5. P.Mahesha and D.S. Vinod, “Vector Quantization and MFCC based Classification of Dysfluencies in Stuttered Speech,” Bonfring International Journal of Man Machine Interface, Vol. 2, No. 3, September 2012.
6. Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar “A Review on Speech Recognition Techniques” International Journal of Computer Applications (0975 – 8887) Volume 10– No.3, November 2010.
7. S.C.Shekokar, Prof. M. B. Mali, “A brief survey of a DCT-Based Speech Enhancement System,” International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013.
8. Titus Felix Furtuna, “Dynamic Programming Algorithm in Speech Recognition,” Revista Informatica Economica nr.2 (46)/2008.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2015

BIOGRAPHY

V.Naveen Kumar was born in Telangana, India in 1983. He is working as Project Engineer at Research and Consultancy Center (RCC) in VNR Vignana Jyothi College of Engineering & Technology (VNRVJIET), Hyderabad, Telangana, India. He completed M.Tech in Embedded systems in 2009 from VNRVJIET and B.Tech in Electronic & Communication Engineering from AZCET, JNTU Hyderabad. He has five years of research experience. His interests include Wireless sensor networks, Embedded Systems, RFID, Microcontrollers and signal processing. He has two patents in wireless stream and eight international journals in various streams.

Dr. Y Padma Sai, works as Lecturer in the Department of ECE in Deccan College of Engineering and Tech, Hyderabad and Later joined as an Assistant Professor in ECE at VNRVJIET in July 1999.Atpresent she is Professor and Head of the department of ECE .Her main objective is to impart quality education and learn New technologies and the scope is to fill gap between industry and academics.

C. Om Prakash received the B.Tech degree in electronics and communication engineering from Sri K S Raju Institute of Technology and sciences, affiliated to Jawaharlal Nehru Technological University Hyderabad, AP, India, in 2012, He has done M.Tech in Embedded systems at VNR Vignana Jyothi Institute of Engineering& Technology, Bachupally, Hyderabad, India.