

Voice Command Based Computer Application Control Using MFCC

Abinayaa B., Arun D., Darshini B., Nataraj C

Department of Embedded Systems Technologies, Sri Ramakrishna College of Engineering, Coimbatore, Tamilnadu, India

ABSTRACT: Speech processing is an efficient application area of digital signal processing. Speech recognition, Speaker recognition, voice command based applications, etc uses speech signals for their processing. In speech processing speaker recognition is an interesting research field and still there are number of problems unresolved. Speaker recognition is determined by the speaker identity. The speaker's identity can be recognized by their voice, which must be familiar to humans and also to machines. This paper speech processing into two components, one is speaker identification and the second is speaker verification. Speaker identification includes training of voice and speaker verification includes testing of the acquired voice with the trained voice. MATLAB software with its associated MFCC is used for speech training and testing and for representing speech signals. In the training phase the input voice will be given. In the testing phase the input voice will be checked and authenticate and opens the corresponding application.

I. INTRODUCTION

Speech recognition research field has become more powerful nowadays. The interactive speech applications are becoming more popular now-a-days in the market. Speech recognition systems allows us to input voice to an application rather than getting input from mouse, keyboard or any other input forms. This paper establishes speech processing in two forms, speech identification and speech verification. MATLAB carries out these two steps with the help of MFCC. MFCC determines the feature extraction and feature matching. Depending upon the application the features of the speech are extracted and matched to open an application. The main application fields of speech processing are security control, telephone banking, shopping through telephone, voice mails, etc...

II. BLOCK DIAGRAM

Speech processing in this paper is carried out through MATLAB software and MFCC technique. Speech training and verification is through MATLAB processing and the operating system accomplishes the applications which have to be processed on. The voice inputs are given as commands to MATLAB, testing and verification are done and corresponding application is opened.

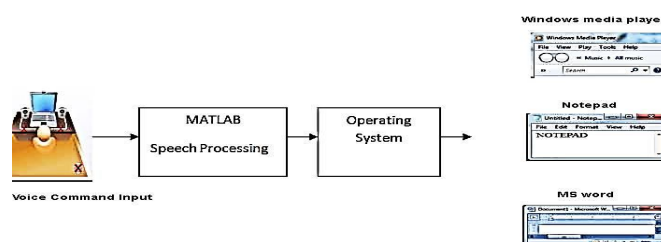


Fig1: Block Diagram

III. SPEAKER RECOGNITION

Speaker recognition is divided into two modules namely speech identification and speech verification. The process which determines the registered speaker's utterance is called Speaker Identification whereas speaker verification is the process which associates accepting and verifying the identity of the registered speaker. The next highest level of the speaker recognition system involves Feature Extraction and Feature Matching. Feature Extraction extracts the features of the voice signal that identifies the speaker, whereas Feature Matching accomplishes the procedure of determining the unknown speaker by comparing the features of the voice input with the stored features of voice in the database.

Speaker recognition application systems have two basic functionality phases they are training phase and testing phase. Training phase is also known as enrolment phase, where each speaker has to register themselves by providing samples of their voice thus forming a reference module. Testing phase matches the input speech signal with the reference module to identify the speaker; it is also known as operational phase.

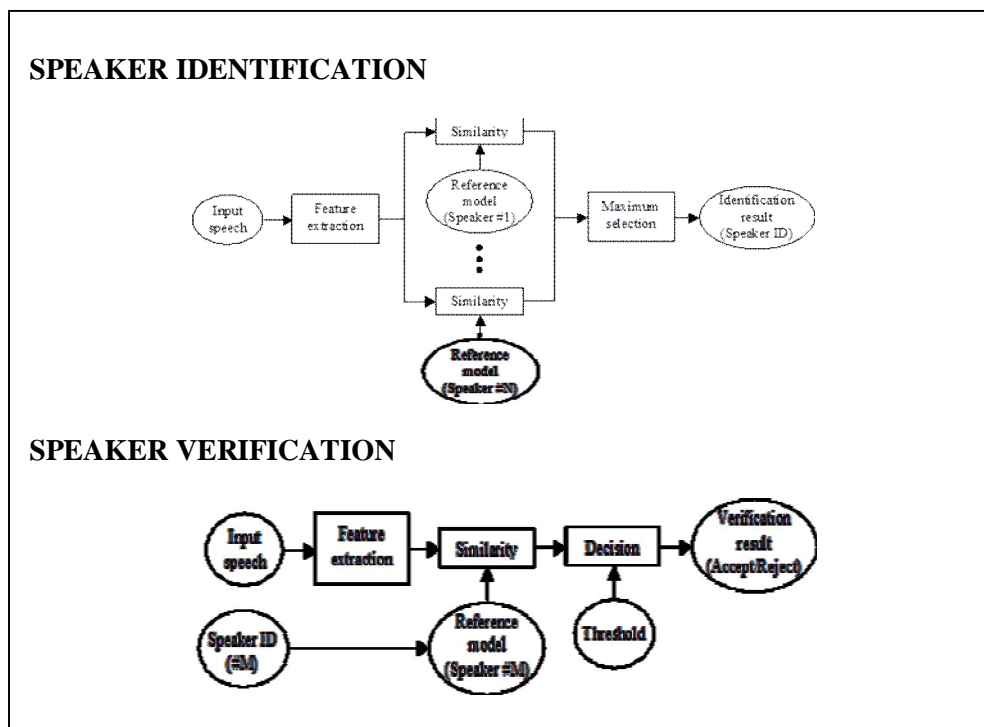


Fig. 2. Speaker Identification and Speaker Verification.

IV. SPEECH FEATURE EXTRACTION

This module converts the speech signal using DSP tools to a particular set of features for future analysis. This type of analysis is considered as front end of signal-processing. We can say the speech signal as slowly varying time signal i.e., quasi-stationary. To characterize the speech signal, short-time spectral analysis is the most common mechanism. MFCC is the most common technique to characterize the speech signals. It is mainly based on the variation of critical bandwidths of the human ears.

a. MEL-FREQUENCY CEPSTRUM COEFFICIENTS PROCESSOR

The sampling rate of about 10000Hz is been chosen to typically record the input speech signal. This sampling frequency has been chosen to minimize the aliasing effects during A/D conversion. The sampled signals are captured at the frequencies upto 5 kHz, which includes most of the sounds generated by the humans. The main functionality of the MFCC processor is to mimic the human ears behavior.

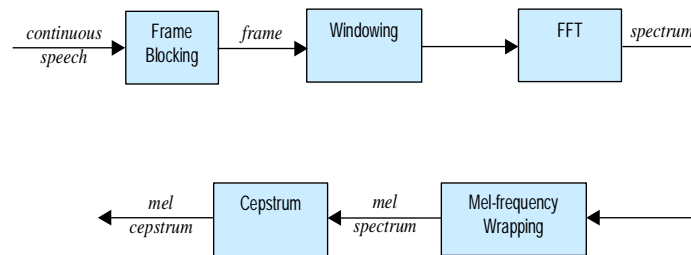


Fig 3., Block Diagram of MFCC Processor

A.1 Frame Blocking

The continuous input speech signal is formed as frames of N samples and adjacent frames being separated by M ($M < N$). The first frame consists of first N samples; the second frame begins M samples after the first frame and so on. This process continues until all the speech is accounted for within one or more frames. Typical values of N and M are $N = 256$ and $M = 100$.

A.2 Windowing

To minimize the signal discontinuities at the beginning and end of each frame, each individual frame undergoes windowing technique. Spectral distortion can be minimized by using the window to make the signal to zero at the beginning and end of the frame. Typically hamming window is used, which is defined as:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1$$

A.3 Fast Fourier Transform

The next step is the Fast Fourier Transform, which converts the frame from time domain into the frequency domain. The FFT is considered as the fast algorithm to implement the Discrete Fourier Transform (DFT), for N samples $\{x_n\}$, as follow:

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

where X_k 's are complex numbers. The result of this step

is referred to as *spectrum*.

A.4 Mel-frequency Wrapping

The psychophysical studies determine the human perception that the frequency of sound of the speech signals does not follow a linear scale. Thus each tone with an frequency (f), measured in Hz, pitch is measured on a scale called the 'mel' scale. The *mel-frequency* scale is considered as a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.

A.5 Cepstrum

Finally we convert the log mel spectrum back to time. The result of this step will be the mel frequency Cepstrum coefficients (MFCC). The Cepstral representation provides a better representation of the spectral properties of the signal. The mel spectrum coefficients are real numbers, so we convert them to the time domain by using the Discrete Cosine Transform (DCT).

V.SPEECH FEATURE MATCHING

Pattern recognition is the topic in science and engineering to be concentrated on. It is considered to be a tedious technique in speech recognition. The goal of this technique is to classify the objects into number of categories or classes. The classes specified here are referred to as individual speakers. The group of patterns or acoustic vectors extracted from the input speech signal uses the techniques used in previous sections to match and verify the patterns. This technique is also referred to as feature matching.

The extracted feature undergoes certain feature matching techniques to classify the speaker. The extracted feature are been tested with the features of the input speech signal which is trained during the training phase. The testing phase tests and validates the features and identifies the speaker.

The feature matching technique used in speaker recognition includes Vector Quantization (VQ), Hidden Markov Modeling (HMM), and Dynamic Time Warping (DTW). In this paper, Vector Quantization approach will be used, since it is easy to implement with high accuracy. Vector Quantization (VQ) is the implementation of mapping vectors from large space of vector to the space with finite number of regions. Each region is considered to be as the cluster and represented by the center called codeword. The collection of code words is called the codebook.

V. SOFTWARE DESCRIPTION

A. MATLAB

In our application we have used MATLAB for voice reorganization. It can be used into three different types of modes such as 'text to speech', 'speech to text' and also as 'voice command recognizer'. We have implemented our application in third mode. In voice command recognizer mode of operation we can add predefined commands. It listens to the command and matches it from the given list. If match occurs it generates an event which corresponds to the match.

a. ALGORITHM

- Step 1: Start the training code.
- Step 2: Enter the training voice
- Step 3: Record the training voice.
- Step 4: Implement the voice into the MFCC tool at the baud of 8000Hz.
- Step 5: Store the MFCC output value.
- Step 6: Start the testing code program.
- Step 7: Enter the testing voice.
- Step 8: Compare the testing and training voice.
- Step 9: If it is equal then open the particular application.

VI. EXPERIMENTAL RESULTS

This figure represents the training phase of the application. MATLAB gets the input as commands with voice input. For accuracy and precision the voice is given as samples for 5 to 10 times.

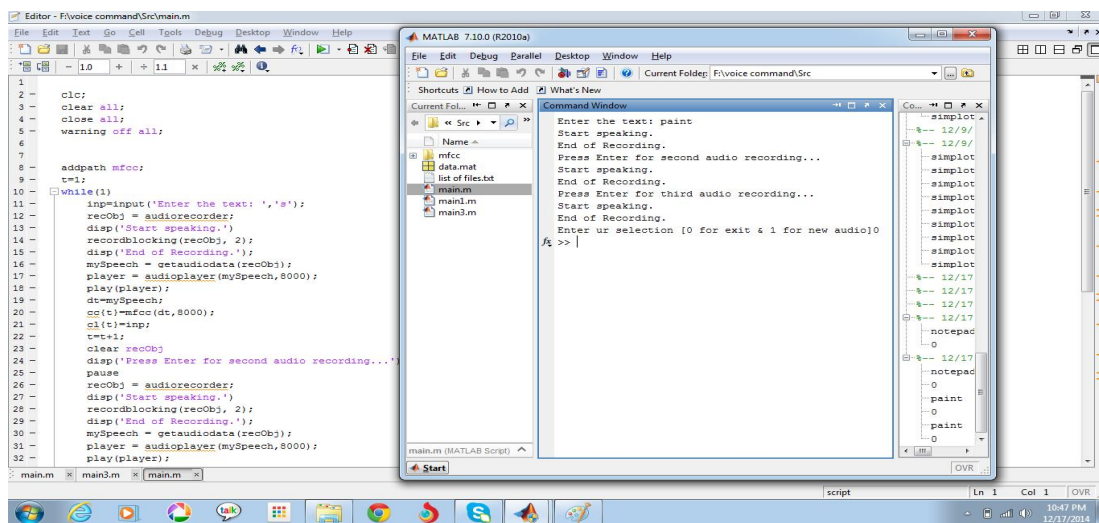


Fig 4., Training Phase

The following figure represents the testing phase and the result obtained after the testing phase. As the result of testing phase, the paint application is been opened.

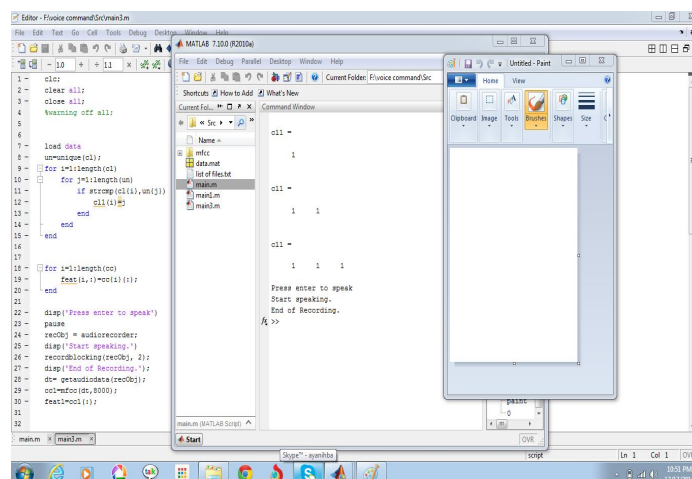
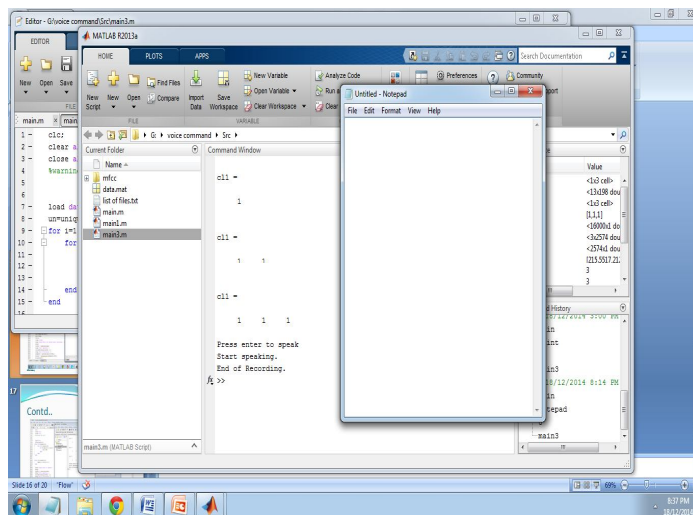
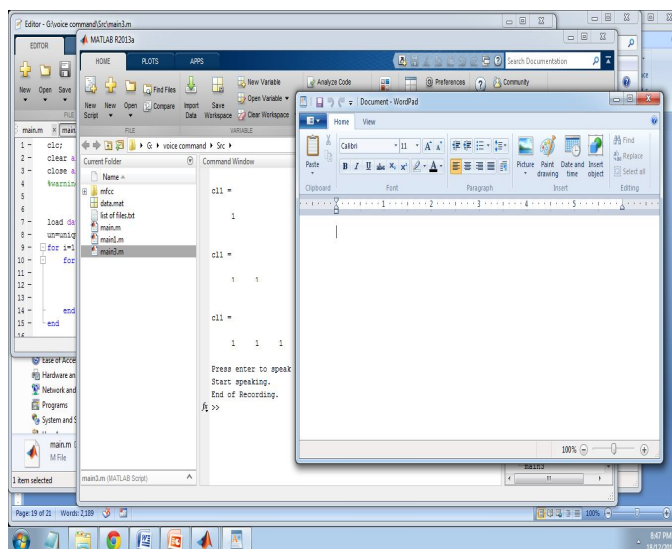


Fig 5., Result of Testing Phase

For Notepad application:



For wordpad application:



VII. CONCLUSION AND FUTURE WORKS

A system for reliable and efficient recognition of voice has been designed and developed. This is very easy application to control the appliances and even an ordinary person can control it, also very much cost effective and easy to design. By using this we can open some of the windows applications such as notepad, word pad, paint etc. This provides us to secure our documents from others i.e., it provides security to our personal and confidential documents. In future we can enhance this application to control and implement real time applications like robotics; voice command based automobile applications, health care appliances, and home automation. Further enhancement will take greater scalable measures in agricultural field, and will help handicapped people to large extent.

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization,

Volume 4, Special Issue 4, April 2015

National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2015)

On 20th & 21st March

Organized by

Sri Krishna College of Engineering & Technology, Kuniyamuthur, Coimbatore-641008, Tamilnadu, India

REFERENCES

- [1] Darshan Mandalia, Pravin Gareta– “Speaker Recognition Using Mfcc And Vector Quantization Model”.
- [2] Aditya Agarwal, Subhayan Banerjee , Nilesh Goel “Design & Implementation Of A Person Authenticating & Commands Following Robot.”
- [3] 39609152-Voice Recognition.Pdf/Www . Troubleshoo T 4 Free . Com / Fyp /
- [4] <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>
- [5] <http://in.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/mfcc.m>
- [6] E. Darren Ellis Department of Computer and Electrical Engineering – University of Tennessee, Knoxville Tennessee 37996 topic on “Design of a Speaker Recognition Code using MATLAB”
- [7] Topic on “Speech Recognition using Digital Signal Processing” ISSN:2277-9477, Volume2, Issue 6.
- [8] Ms. Arundhati S. Mehendale and Mrs. M.R. Dixit "Speaker Identification" Signals and Image processing: An International Journal (SIPIJ) Vol. 2, No. 2, June 2011.
- [9] <http://labrosa.ee.columbia.edu/~dpwe/papers/MakhRG85-vq.pdf>
- [10] “Controlling of Device through Voice Recognition Using Matlab1” ISSN NO: 2250-3536 VOLUME 2, ISSUE 2, MARCH 2012 http://www.ijater.com/Files/fd54f0ee-1aae-4e88-ab81-b89f32ac57c6_IJATER_03_35.pdf
- [11] <http://www.fon.hum.uva.nl/david/sspbook/sspbook.pdf>