

HANDS FREE COMPUTER CONTROL

Pooja Antony¹ & Prof. Sunny Joseph²

¹Department of Electronics, Mar Athanasius College of Engineering, Kothamangalam,

A .P.J Abdul Kalam Technological University, Kerala, India

²Associate Prof., Department of Electronics & Communication,
Mar Athanasius College of Engineering, Kothamangalam

ABSTRACT

For modern human computer interaction, traditional method based of using keyboard and mouse do not prove sufficient. Especially when thinking of physically handicapped persons, we must try to find the easiest and most comfortable method for human system interaction. The good method for communication process going from man to machine is speech communication. In case of control system, speech communication offers free hand manipulation, which is sometimes very important. In modern era, mouse control has become an important part human computer interaction, which is difficult for physically disabled persons. This paper 'Hands free computer control' presents two systems called as vocal mouse and virtual keyboard. The device vocal mouse will allow the users to work on continuous motion control over computer screen including the virtual keyboard created as GUI. This includes commands consists of nine mouse controlling commands, low level acoustic features are extracted in real time using MFCC. Pattern recognition is performed using minimum feature distance technique. Vocal mouse can be used by users without extensive training which is the main advantage of the system.

KEYWORDS

MFCC, GUI, k-mean clustering, Speech recognition

1. INTRODUCTION

Voice input has a number of potential benefits, especially for physically disabled people. People who are suffering from physical disabilities cannot use standard input devices like keyboard and mouse to access the system, so they have only one option for gaining access to the computer i.e. hands-free input methods[2].

The project "Hands Free Computer Control" mainly presents two systems, Vocal Mouse and Virtual Keyboard. Vocal Mouse is an enhanced voice-based interaction will give benefit to physically handicapped people who find themselves in impairing situations like driving or interacting with computer, hands-free interaction can be more suitable than traditional manual input devices.

This system was reliable and speaker independent. Operator training was not required. Any 32-bit Windows software, can access this voice control system. Standard programming languages such as Borland or Microsoft C/C++ and Visual Basic (VB) and commercial packages such as Lotus Notes and Microsoft Word support this system. In this system, the user is allowed to enter data and to control the software flow by voice command or from the keyboard or mouse.

Main objective of designing Vocal Mouse is to make interactions with existing computer applications possible. This objective put more emphasis on the practical needs of physically challenged people who cannot use existing computer and applications, which have been designed with keyboard and mouse inputs. However, Vocal Mouse cannot not provide the ideal solution from the interaction design perspective. Even then it will give benefit to people who are physically challenged more than the alternative of not being able to access the functionality of computers at all. Another objective of the proposed system is to provide interactions with existing computer applications more effective for any users. The main aim of this system is to use voice input can as an additional input to augment the standard keyboard and mouse interaction for greater control i.e. giving sound input parallel to the keyboard and mouse. The other motivation is to design a new application environments optimized for voice-based control.

2. PROPOSED SYSTEM

The speech command or input can be determined by power of audio signal which can be taken by the help of microphones being connected to the computer itself. Using MATLAB the speech signal is sampled at a rate of 8000 samples/sec according to nyquist criteria i.e. $F=2*f_m$ Where F =Sampling Frequency, f_m = maximum component of frequency being present in speech signal. Then the sampled signal is filtered off by using band pass filter lying in the range of 300 Hz-4000 Hz, which filters all the audio signals lying below 300 Hz of frequency range. Moreover, it includes the algorithm for the creation of speech parameters which can be achieved by calculating the power of each sampled signals respectively.

Initially, a database is created in which same command is taken many times and thus the average of these will represent a speech template which can be stored in a library. Now the command speech signal that are being received by microphone are then compared with the speech signal being stored in library according to Euclidean's Distance. Thus, the command will be detected by the system and it will performs the operation accordingly & if it does not match, the system will not perform any operations. The general block diagram of the system is shown in figure2.1.[2]

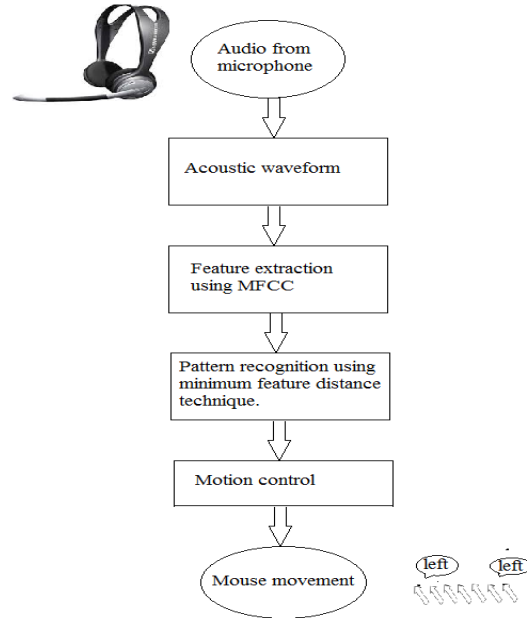


Fig 2.1 Block diagram to represent the working of hands free computer control

This system truly based on speech recognition which can be achieved by the use of Mel Frequency Cepstrum Coefficients which will process the input audio signal and further will perform the speech recognition. This above task can be performed by using MATLAB Programming. In the above diagram, when the input signal i.e. the speech signal is provided the Speech Extraction process simply extracts all the signal components such as pitch, frequency etc. which are needed at the end for the recognition of speech. Then, the Speech Matching Block will match all the properties that are extracted before with the set of audio command signals being stored as database.

The speech recognition system consists of two phases. The first one is referred to the training phase while the second one is referred to as the testing phase.

2.1. BACKGROUND ON SPEECH FEATURE EXTRACTION[1]

Methodology of constructing the proposed system will consists of different modules. Each module uses different methods and algorithms to perform its tasks. After a particular module completes its task, the output will be given as input for the next module.

- 1:- Acoustic signal processing
- 2:- Pattern recognition
- 3:- Motion control.

2.1.1 Signal Processing

Speech feature extraction is a fundamental requirement of speech recognition system; it is the mathematical representation of the speech signal. In a human speech recognition system, the main

aim is to classify the source files using a reliable representation that reflects the difference between utterances of audio signals.

2.1.1.1 Pre-processing

Pre-processing is the fundamental signal processing method applied before extracting features from speech signal, for enhancement of the performance of feature extraction algorithms. Commonly used pre-processing techniques contain DC component removal, pre-emphasis filtering, and amplitude normalization.

2.1.1.2 DC Component Removal

The speech signal often has a constant component, i.e. a non-zero mean known as the DC component of the signal. This is typically due to DC bias within the recording instruments. The DC component in the input signal can be easily removed by subtracting the mean value from all samples within an utterance.

2.1.1.3 Pre-emphasis Filtering

A pre-emphasis filter consists the dynamic range of the speech signal's power spectrum by flattening the spectral tilt. The filter is in form of

$$P(z) = 1 - az^{-1}$$

Where a ranges between 0.9 and 1.0

In speech processing, the glottal signal can be modelled by a two-pole filter with both poles close to the unit circle. However, the lip radiation characteristics models its single zero near to $z=1$, which tends to cancel the effect of one of the glottal pole. In addition, the spectral slope of a human speech spectrum is usually negative since the energy is concentrated in low frequency. Thus, a pre-emphasis filter is introduced before applying feature algorithms to increase the relative energy of the high-frequency spectrum.

2.1.1.4 Amplitude Normalization

Recorded signals often have varying energy levels due to speaker's volume and the distance of microphone. Amplitude Normalization can cancel the inconsistent energy level between signals, and thus can enhance the performance of the energy-related features. There are several methods used to normalize a signal's amplitude. One of them is achieved by the point-by-point division of the signal by its maximum absolute value, so that the dynamic range of the signal is constrained between -1.0 and +1.0.

2.1.1.5 Mel Frequency Cepstral Coefficient (MFCC) [1]

The first step in speech recognition system is to extract features i.e. identify the components of the audio signal that are good for identifying the linguistic content in the signal and discarding all the other things which carries information like background noise, emotion etc.

Mel Frequency Cepstral Coefficients (MFCCs) is one of the feature widely used in automatic speech and speaker recognition. Prior to the introduction of MFCCs, Linear Prediction

Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) were used for automatic speech recognition (ASR).

Steps InCalculating MFCC:

1. Frame the signal into short frames.
2. Calculate the periodogram estimate of the power spectrum for each frame.
3. Apply the Mel filter bank to the power spectra; sum the energy in each filter.
4. Take the logarithm of all filter bank energies.
5. Take the DCT of the log filter bank energies.
6. Keep DCT coefficients 2-13, discard the rest.

The Mel scale relates perceived frequency, or pitch, of a pure tone to its actual measured frequency.

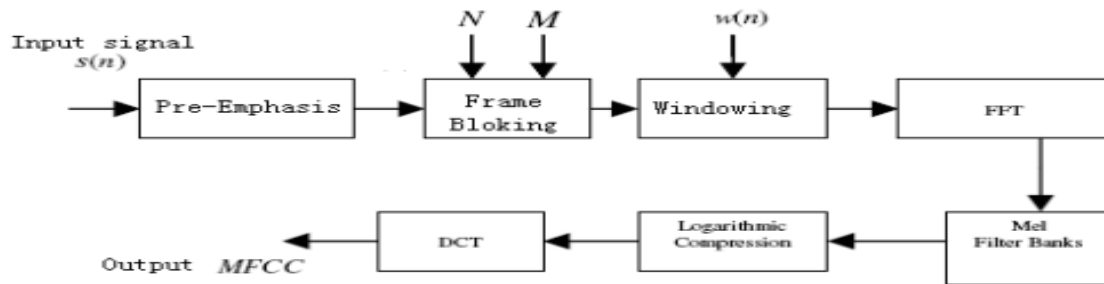


Fig 3.1 Mel Frequency Cepstral Coefficients[1]

2.1.1.6 Windowing Functions

Speech is a non-stationary signal. Thus a window function is applied in the pre-processing stage to divide the speech signal into small segments to approximate a stationary signal,. There are many window functions are available. The proposed system is designed using *Hamming window*, defined as

$$w(n) = \begin{cases} (1 - \alpha) - \alpha \cos \left[2\pi n / (N - 1) \right] & 0 \leq n \leq N - 1 \\ 0 & n = \text{else} \end{cases}$$

Where N is the window length and $\alpha = .46$

2.1.1.7 K-Mean Clustering

K-means clustering is a vector quantization method, originally from signal processing, that is popular for cluster analysis. k-means clustering algorithm is used to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Let $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ be the set of observations, where each observation is a d -dimensional real vector, k -means clustering partitions the n observations into k ($k \leq n$) sets $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$ so that it minimize the within-cluster sum of squares (WCSS).

2.1.2 Pattern Recognition[2]

This module uses the various audio features to extract desired parameters. Approach used for pattern recognition is:-

Minimum Feature Distance Technique

This technique is based on calculating distances between the spoken word and each word in the library.

- 1) $D = \text{features of spoken word} - \text{features of word stored in library at training time.}$
- 2) Sum up all the corresponding differences.
- 3) Take the square root of the total calculated difference.
- 4) For each command in library perform the above step 1, 2 and 3 calculations
- 5) Above steps will result in nine feature distance values
- 6) Print the word with minimum feature distance.
- 7) The result will correspond to word spoken by user

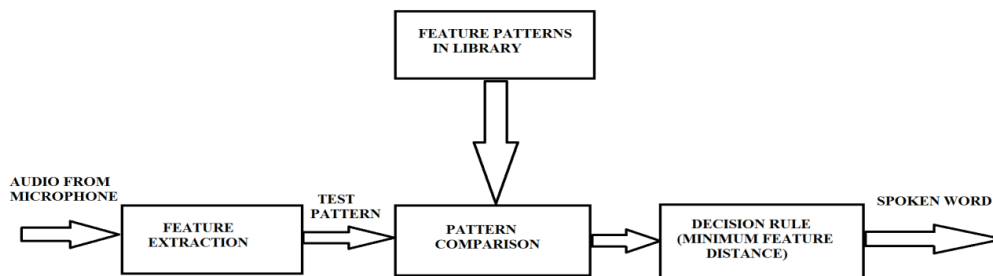


Fig. 3.3 Pattern Recognition[2]

2.2. SPEECH RECOGNITION

Speech recognition refers to the process which finds the distance or probability between the vector sequence of the unknown speech feature and each input speech sample, when matching the unknown speech features with different training patterns. The system digitized analog signals according to Nyquist sampling frequency. After pre-processing the input speech signal by using several algorithms, Mel-frequency cepstral coefficient (MFCC) is calculated. The MFCC analysis consists of four steps.

- a) Perform a fast Fourier transform on the input speech signal.
- b) Model the frequency axis by the Mel-scale. The Mel-frequency melf can be computed from the frequency f
- c) Calculate the amplitude spectrum of the triangle filter to the filtered signal output
- d) Perform the discrete cosine transform on the logarithm of the filter-bank energies and append first order differentials.

Then Cluster the features of the training speech samples using K-means clustering algorithm. Fig. 2.2 shows that when recognizing, the system calculates the feature of input speech and determines

its cluster group k at first. In the case of appropriate cluster group the system will save a considerable amount of computation. Then consider the test file, input the test command through microphone if the input is not strong enough then displays the message ‘TRY AGAIN’. This is to avoid noise signals in the surroundings. Then after pre-processing perform MFCC calculation to extract MFCC features. Calculate the distances between the test speech samples and the cluster centres. For each sample, the minimum distance determines its Euclidian distance. Check whether the target group contains the training sample. If included, corresponding command displayed on the screen and the mouse pointer moves up , down or in specified direction

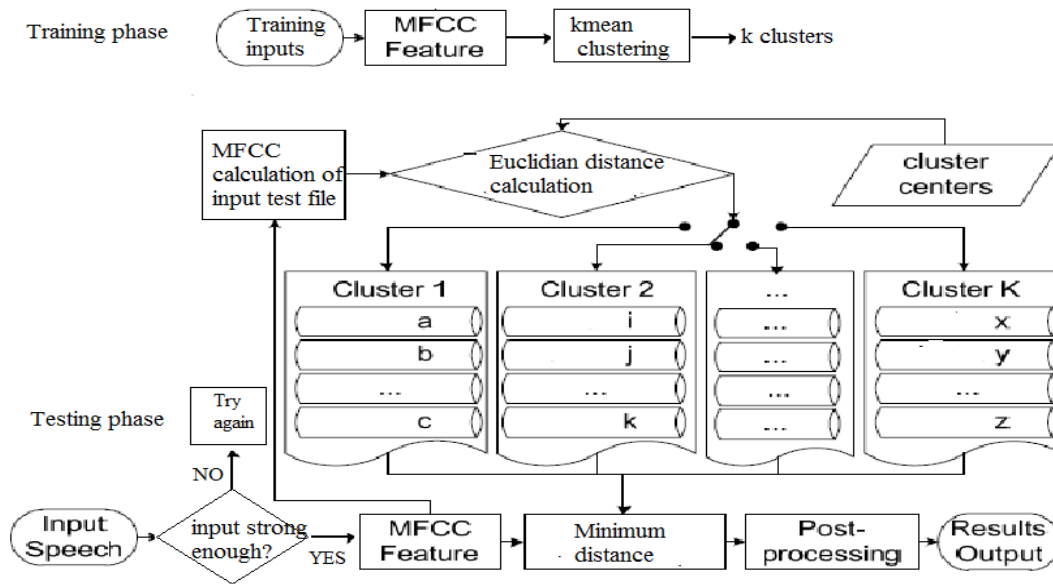


Fig.2.2 flowchart for training phase and testing phase

2.3. GRAPHICAL USER INTERFACE

“Hands free computer control” allows the user to continuously control various applications running on the system using vocal mouse and virtual keyboard. A set of operating modes which include, database recording and recognition for sound, and virtual keyboard for documentation are integrated under the same programming platform. This requires a set of graphical user interfaces (GUI).The main GUI is provided to allow the users to select the type of function:-

- [1] Recording of speech commands
- [2] Adding the speech commands to database
- [3] Testing the input command from the microphone with the database.
- [4] Virtual keyboard for various functions like creating and editing documents, and other keyboard applications
- [5] Pop up menu for pointer location selection.

[6] Exit from main GUI

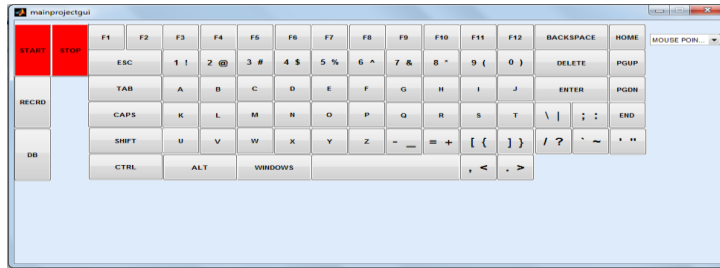


Fig2.3 Main GUI

2.3.1 Creation Of Database Using GUI



Fig. 2.4GUI for Database creation

If user selects 'START' option another GUI pops up saying to record the sound after pressing OK. And when we press OK two more dialog boxes appear on the screen saying to give a name to the .wav file and the name of the character to be stored in the database. Then select 'DATABASE' for adding the sound file to database.

2.3.2 Testing

The sound recognition and matching starts with 'TEST' option. After clicking on this option the user will have a small time gap for making the sound. The system process the sound based on MFCC calculation and gesture corresponding to best match is displayed.



Fig 2.5 command window(MATLAB)

3. IMPLEMENTATION TOOL AND RESULTS

The proposed system can be implemented using MATLAB. To do work on speech recognition, users can take the help of VOICEBOX. There is no need of any other expensive, bulky hardware.

The only thing that is needed for the system is good quality microphone.

A. Training Phase

Vocal Mouse project starts with training phase. Get nine commands (up, down, left, right, click, doubleclick, right click, start and stop) from microphone and compute their features and to create database. Initially, save the calculated features in a feature matrix and then store them in some other file. User is given 1 second to say each word. User can press START button and say the specified word in 1 second. Features of all the spoken words are stored in a feature matrix. Matrix is a 2-D matrix with 9 rows (one for each word) and 17 columns (17 features are extracted). Contents of this matrix can be loaded to library data file for further reference. Here is the completion of training phase.

B. Testing Phase

After clicking 'TEST' button on the GUI, the Testing phase started. During the testing phase, User can continuously say any word. Features of the spoken words are computed and these features thus extracted are compared with feature patterns already stored during training time. For this comparison, approach is used is minimum feature distance. Vocal Mouse system will compute resulting spoken word on the basis of minimum distance. Output will be displayed on the console window of MATLAB giving spoken word by the user. After the command recognition, system will perform task as per the requirement of the user.

4. FUTURE SCOPE

One area of enhancement of the proposed system is in the user interface for supporting self-diagnosis of recognition issues. The user can be provided with a testing mode to test whether the system is responding properly or not. More information should be conveyed to the user that would allow them to troubleshoot when the system fails to recognize certain inputs. The main obstacle and sources of frustration that a user faces is false signal that occur while operating the system. False positive means that system recognize event when the user did not intend to vocalize. This will happen when the user forgets that the system is processing vocal input and begins speaking or making some sounds, typically when the users attention is away from the user interface with the system, or when the system picks some background noise and incorrectly recognizes it as some valid vocalization. In these cases, the user cannot not realize that the system has processed the false positive events later, e.g., when the user returns his attention back to the system, at which point the user become confused what had happened, and possibly quite frustrated about not knowing what to undo.

In such a situation, quick method is required to disable current operation. A possibility for future work for addressing this issue is the use of various external contexts such as the users head posture and gaze to disable voice input when the user is likely disengaged from the system interface. Current method of Hands free Computer control uses only standard spoken words such as up, left, right, down, click etc is inefficient for performing continuous tasks. Therefore the system can be improved by including commands given as words or regular sounds such as continuous speech.

5. CONCLUSION

The Hands Free Computer Control is a system that enables the user to continuously control the mouse cursor using their voice. This uses a technique minimum feature distance algorithm for pattern recognition. The system provides more interactive and easy to user interface. Hands free Computer control is mainly designed for physically handicapped persons and mainly presents two systems, Vocal Mouse(VM) and Virtual Keyboard. Now, VM will allow users to work on continuous motion control over the computer screen including the virtual keyboard created as GUI. This includes commands consist of nine mouse controlling command. The system is based on the recognition of the commands and vocal sounds which is a very robust and accurate method as compared to the recognition of words using conventional speech recognition systems.

ACKNOWLEDGEMENT

The author would like to thank Mrs. Jibi John, Associate Professor, Govt. Model Engineering College, Cochin, for the guidance and support.

REFERENCES

- [1] Aseemsaxena, Amitkumarsinha, "speech recognition using matlab", International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058 Volume- 1, Issue- 2, Nov-2013
- [2] "Mouse Movement using Speech and Non-Speech Characteristics of Human Voice", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-5, June 2012.
- [3] "To Design Voice Control Keyboard System using Speech Application Programming Interface", IJCSI, Vol. 7, Issue 6, November 2010, ISSN : 1694-0814.
SelinaChu, ShrikanthNarayana, and Jay Kuo, " Sound Recognition With Time and Frequency Audio Features" IEEE Transactions audio,speech,and language processing, VOL.17,NO.6,August2009.
- [4] M Abdeen, H Moshammad and M C E Yagoub, "An Architecture for Multi-Lingual Hands Free Desktop Control System for PC Windows", Niagara Falls, Canada:IEEE , 2008

Authors

Pooja Antony, received BTech degree in Electronics and Communication Engineering from Cochin University Of Science And Technology, Kerala in 2015. Currently pursuing MTech degree in VLSI and Embedded Systems from APJ Abdul Kalam Technological University, Kerala.



Prof. Sunny Joseph is the Head of the Department of Electronics and Communication Engineering in Mar Athanasius College of Engineering, Kothamangalam, Kerala, India. He has a teaching experience of 29 years. He received his B.E from Bangalore University and M.Tech degree from Kerala University. His current research focus is in the area of VHDL, high speed digital design

