ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Review on Computer Control with Voice Command (MFCC) using Ad-hoc Network

Pooja, Mukesh Kumar

M.Tech (pursuing), Dept. of Electronics & Communication Engineering, Shri. Ram College of Engineering and Management, Palwal, Haryana under the Affiliation of Maharshi Dayanand University at Rohtak, Haryana, India

Assistant Professor, Dept. of Electronics & Communication Engineering, Shri Ram College of Engineering and Management, Palwal, Haryana under the Affiliation of Maharshi Dayanand University at Rohtak, Haryana, India

ABSTRACT: It has continually been a dream of soul to form machines that behave like humans. Recognizing the speech and responding consequently is a vital a part of this dream. With the enhancements of the technology and researches on artificial intelligent, this dream comes true comparatively, during this theme it's aimed to form a contribution to the present dream. Dominant the machines and atmosphere with speech makes human life easier and more leisurely. This proposed scheme may be a straightforward implementation of this approach. The machine is controlled by voice commands using adhoc-network . Voice command is taken through a movable devices like mobile phones with ad-hoc network, processed in laptop and sent to the operation system and eventually the mechanism acts consequently. Speech is that the most vital method of communication for individuals. Inculcating the speech as interface for processes became a lot of necessary with the enhancements of artificial intelligent. The options were extracted with The Mel Frequency Cepstral Coefficients (MFCC) algorithms and that they were recognized by the assistance of Artificial Neural Networks. Finally the comments were regenerating the shape within which the software system will acknowledge and act consequently.

KEYWORDS: 802.11 tethered network, Mel Frequency Cepstral Coefficients (MFCC), Hidden Markov Model (HMM), Dynamic Time Wrapping (DTW), Differential pulse code modulation (DPCM), Linear Predictive Codes (LPC).

I. INTRODUCTION

Speech is that the most used means of communication for peoples. We have a tendency to born with the talents of speaking, learn it simply throughout our time of life and principally communicate with one another with speech throughout our lives. By the developments of communication technologies within the last era, speech starts to be a very important interface for several systems, rather than victimization advanced completely different interfaces, speech is less complicated to speak with computers. during this project, it's aimed to regulate a machine with speech commands. The machine is ready to acknowledge spoken commands to manoeuvres properly, to provide a direction to machine, initial the voice command is send to the pc employing the itinerant however, it acknowledges the command by speech recognition system. consequently pc converts the voice command to direction command that predefined and recognizable by software. there's conjointly the windows utility referred to as remote desktop which supplies various geographically areas the flexibility to remotely hook up with a computer/PC within the network, once obtaining connected to the pc the screen of the pc seems on the machine from wherever folks square measure connecting. once the association is self-made folks will management the computer as if it's their own computer and folks square measure dominant it with their keyboard and mouse. however this windows utility is totally completely different from my project because it don't connect the machine however still will management the resources to lock and unlock them. this manner is saves the process power of each the server and therefore the shopper computers, so rushing up the method and conjointly provides movability via a hand-held phone. within the projected resolution it'll write associate application in java with to completely different parts as server part and controlling system with increased accuracy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Below diagram depicts the workflow the proposed scheme are research work as figure 1:

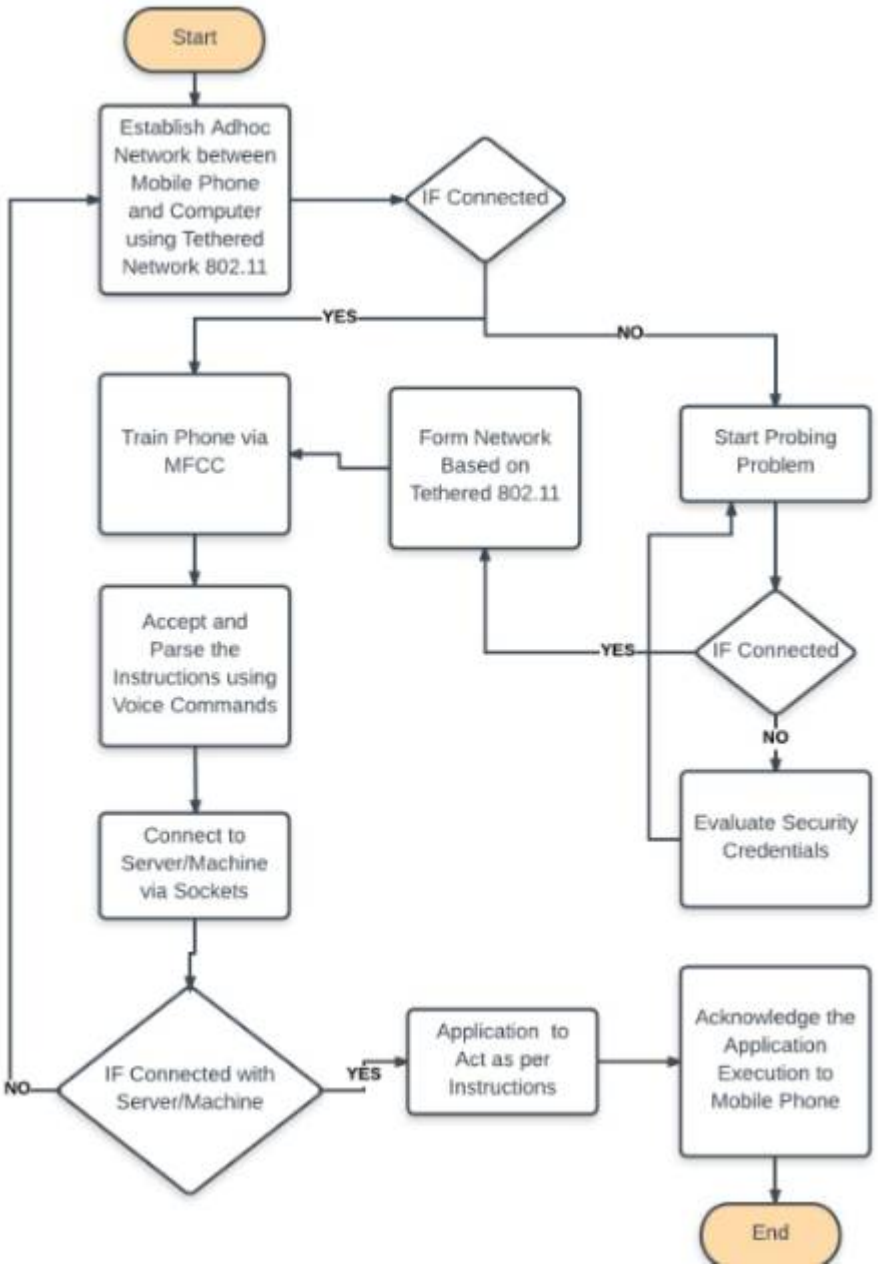


Figure 1: Workflow diagram of proposed scheme forming ad-hoc network using MFCC vide mobile phone and transmitting data using Sockets to the computer/machine/laptop

Feature Extraction : The signal, as first captured by the microphone, contains information in a form not suitable for pattern recognition. However, it can be represented by a limited set of features relevant for the task. These features more closely describe the variability of the phonemes (such as vowels and consonants) that constitute each word. There are different techniques to extract the required features such as DPCM, LPC, MFCC, etc. LPC and MFCC are most successful future extraction techniques and mainly one of these techniques is used in the speech recognition projects [2,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

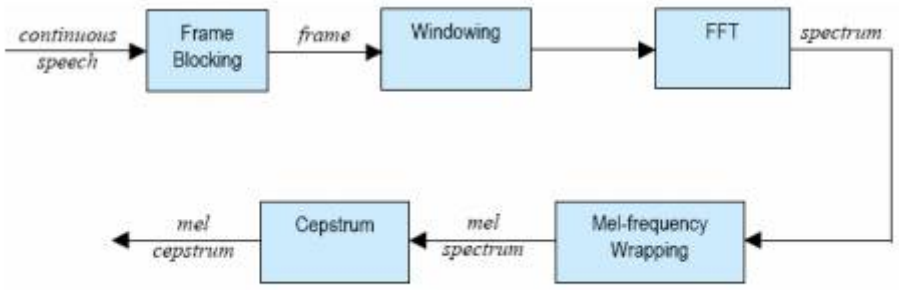
3, 4, 5]. As a comparison of these two techniques, Sahar E. Bou-Ghazale and John H. L. Hansen show these results in their project "A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress"[4] :

Table 1 Recognition performance based on feature extraction techniques [4]

Techniques	Speaking Styles				Overall Recognition
	Neutral	Angry	Load	Lombard	
LPC	%61.65	%37.78	%43.89	%49.44	%48.19
MFCC	%83.52	%58.15	%63.89	%72.22	%69.45

Table 1: Linear predictive coding and Mel-frequency cepstral coefficients Speaking Style Analysis under Stress

Mel Frequency Cepstral Coefficients The speech input is typically recorded at a sampling rate above 10000 Hz. This sampling frequency is chosen to minimize the effects of aliasing in the analogto-digital conversion. These sampled signals can capture all frequencies up to 5 kHz, 9 which cover most energy of sounds that are generated by humans. The main purpose of the MFCC processor is to mimic the behavior of the human ears. In addition, rather than the speech waveforms themselves, MFCC's are shown to be less susceptible to mentioned variations. It is shown below the block diagram of MFCC process [11]:



Block diagram of MFCC Process [11]

Figure 2 : MFCC Work-flow diagram depicting the mel-frequency wrapping and cepstrum capturing.

Recognizition: The most important part of the system is recognizing the phonemes, groups of phonemes and words or utterances. This stage can be achieved by many processes such as GMM (Gaussian Mixture Model), DTW (Dynamic Time Warping), HMM (Hidden Markov Model), NNs (Neural Networks), expert systems and combinations of techniques. Although DTW is an old technique, it is already used for projects which are not very complex such as speaker recognition projects. The most common used techniques are HMM and NNs. These techniques are both very successful in speech recognition. For the speech recognition projects with big vocabulary, HMM is better than NN but for small or medium vocabularies both techniques give satisfactory results [6, 7, 8]. And for more challenging projects such as spontaneous speech recognition with highly noise level these two techniques are used together as Hybrid HMM/NN. As a comparison of these two techniques, Raymond Low and Roberto Togneri show these results [6]:

Speech Recognition performance based on recognizer type[6]

Techniques	Digits	Alphabet	Confusable /e/ set
HMM	%96.5	%82.0	%81.2
NN	%94.1	%88.6	%83.0

Table 2:Comparative analysis of Hidden Markov Model and Neural Network using Hybrid Technique

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Neural Network: Connectionism, or the study of artificial neural networks, was initially inspired by neurobiology, but it has since become a very interdisciplinary field, spanning computer science, electrical engineering, mathematics, physics, psychology, and linguistics as well. Some researchers are still studying the neurophysiology of the human brain, but much attention is now being focused on the general properties of neural computation, using simplified neural models. These properties include:

Trainability: Networks can be taught to form associations between any input and output patterns. This can be used, for example, to teach the network to classify speech patterns into phoneme categories.

Generalization: Networks don't just memorize the training data; rather, they learn the underlying patterns, so they can generalize from the training data to new examples. This is essential in speech recognition, because acoustical patterns are never exactly the same.

Nonlinearity: Networks can compute nonlinear, nonparametric functions of their input, enabling them to perform arbitrarily complex transformations of data. This is useful since speech is a highly nonlinear process.

Robustness: Networks are tolerant of both physical damage and noisy data; in fact noisy data can help the networks to form better generalizations. This is a valuable feature, because speech patterns are notoriously noisy.

Uniformity: Networks offer a uniform computational paradigm which can easily integrate constraints from different types of inputs. This makes it easy to use both basic and differential speech inputs, for example, or to combine acoustic and visual cues in a multimodal system.

Parallelism: Networks are highly parallel in nature, so they are well-suited to implementations on massively parallel computers. This will ultimately permit very fast processing of speech or other data. There are many types of connectionist models, with different architectures, training procedures, and applications, but they are all based on some common principles. An artificial neural network consists of a potentially large number of simple processing elements (called units, nodes, or neurons), which influence each other's behaviour via a network of excitatory or inhibitory weights. Each unit simply computes a nonlinear weighted sum of its inputs, and broadcasts the result over its outgoing connections to other units. A training set consists of patterns of values that are assigned to designated input and/or output units. As patterns are presented from the training set, a learning rule modifies the strengths of the weights so that the network gradually learns the training set. This basic paradigm can be fleshed out in many different ways, so that different types of networks can learn to compute implicit functions from input to output vectors, or automatically cluster input data, or generate compact representations of data, or provide content-addressable memory and perform pattern completion [2]. There are many different types of Neural Networks such as MLP (Multi-layer Perception), PNN (Probabilistic Neural Network), RBF (Radial Based Function), RNN 18 (Recurrent Neural Network), GRNN (Generalized Regression Neural Network), etc. In a comparison of PNN, GRNN and RBF, Bülent Bolat and Üna1 Küçük from Yıldız.

II. LITERATURE REVIEW

Speech recognition is a very important field that can be used in many applications such as banking, and transaction over telephone network database access service, voice email, investigations, and management. In this paper, an approach for recognition isolated Arabic words is presented. Discrete Wavelet Transform (DWT) from type Haar Wavelet with (third and fourth levels) and Magnitude is used in feature extraction stage and Genetic Algorithm (GA) is used in classification stage. The results showed that the recognition rate in third level was 90% and fourth level was 87.5%. Speech Recognition.[1]

Speech is the most important way of communication for people. Using the speech as interface for processes became more important with the improvements of artificial intelligent. In this project it is implemented to control a robot with speech comment. Speech commends were taken to the computer by microphone, the features were extracted with The Mel Frequency Cepstral Coefficients algorithms and they were recognized by the help of Artificial Neural Networks. Finally the comments were converted the form in which the robot can recognize and move accordingly.[2]

It is well known that the performance of speech recognition algorithms degrade in the presence of adverse environments where a speaker is under stress, emotion, or Lombard effect. This study evaluates the effectiveness of

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

traditional features in recognition of speech under stress and formulates new features which are shown to improve stressed speech recognition. The focus is on formulating robust features which are less dependent on the speaking conditions rather than applying compensation or adaptation techniques. The stressed speaking styles considered are simulated angry and loud, Lombard effect speech, and noisy actual stressed speech from the SUSAS database which is available on CD-ROM through the NATO IST/TG-01 research group and LDC1 . In addition, this study investigates the immunity of linear prediction power spectrum and fast Fourier transform power spectrum to the presence of stress. Our results show that unlike fast Fourier transform's (FFT) immunity to noise, the linear prediction power spectrum is more immune than FFT to stress as well as to a combination of a noisy and stressful environment. Finally, the effect of various parameter processing such as fixed versus variable pre-emphasis, filtering, and fixed versus cepstral mean normalization are studied. Two alternative frequency partitioning methods are proposed and compared with traditional mel-frequency cepstral coefficients (MFCC) features for stressed speech recognition. It is shown that the alternate filter bank frequency partitions are more effective for recognition of speech under both simulated and actual stressed conditions.[4]

III. PROPOSED WORK

Despite many years of research, Speech Recognition remains an active area of research in Artificial Intelligence. Currently, the most common commercial application of this technology on mobile devices uses a wireless client – server approach to meet the computational and memory demands of the speech recognition process. Unfortunately, such an approach is unlikely to remain viable when fully applied over the approximately 7.22 Billion mobile phones currently in circulation. In this thesis we present an On – Device Mobile Speech recognition system. Such a system has the potential to completely eliminate the wireless client-server bottleneck vide you can control your machine or laptops For the Voice Activity Detection part of this work, this thesis presents novel algorithms used to detect speech activity within an audio signal. The algorithm is based under the MFCC is augmented in scheme. This algorithm uses the frames within the speech signal with the minimum and maximum standard deviation, as candidates for a linear cross correlation against the rest of the frames within the audio signal. This novel application of the linear cross correlation technique to cepstral coefficients feature vectors provides a fast computation method for use on the mobile platform; as shown by the results presented in this thesis or scheme the proposed scheme is as under:

Algorithm Workflow :-

```
For every frame Fn
        If Fn > Max
          Max = Fn
        End
T = Max x ST
        For every Frame Fn
                If Fn > T
                        Insert Fn into VC
                 Else if
                        Fn < T
                Insert Fn into UC
End Frames
```

- Frames with speech VC
- Frames without speech UC
- Fn = Raw Short time energy of the nth frame
- Max = Value of frame with the highest Cluster value
- T = Threshold
- ST = Selected Threshold (chosen percentage of Max)
- VC = Voice Cluster
- UC = Unvoiced Cluster

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The algorithm initially determines the threshold T, which is a percentage of the maximum value of the frame with the maximum cluster measure. This is determined by first looping through the entire frames and selecting the frame with the maximum cluster value. Then a chosen percentage measure ST is used to determine the threshold T for that audio signal. The chosen percentage measure could be determined by the user. Experiments were conducted to determine the optimum generalised threshold to be used for any input signal. After the threshold T is selected, another variable called frame distance is used to separate the frames in the voice cluster (VC) in order to group them into their respective digits. The second determinant variable called the frame distance is used only after determining the threshold value and effectively clustering the different frames into voiced frames and unvoiced frames.

IV. CONCLUSION

With the above proposed scheme we can achieve controlling computer/laptop from ad-hoc network using mobile phone, whereas, on the part of the computer or laptop under the scheme we will develop the automation service which will accept the connection using 802.11 ad-hoc tether network and bind with computer/laptop which will listen to the requests using sockets, thereinafter under the scheme will develop the mobile app android or iOS based to accept the voice instructions using MFCC and deliver the same to computer/laptop via ad-hoc network. Consequently, over the computer/laptop the listener via socket accepts the instruction and act as per automation workflow/automation defined.

REFERENCES

1. Stephen Cook, Speech Recognition HOWTO
2. Fatih AYDINOĞLU, Robot Control System with Voice Command Recognition, Yıldız Technical University, Department of Computer Engineering
3. P. M. Grant, Speech recognition techniques
4. Sahar E. Bou-Ghazale, Member, IEEE, and John H. L. Hansen, Senior Member, IEEE, A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress
5. M. Chetouani, B. Gas, J.L. Zarader, C. Chavy, Neural Predictive Coding for Speech Discriminant Feature Extraction : The DFE-NPC, Laboratoire des Instruments et Systèmes d'Ile de France
6. Raymond Low and Roberto Togneri, Speech Recognition Using the Probabilistic Neural Network, The University of Western Australia, Department of Electrical and Electronic Engineering
7. M. M. El Choubassi, H. E. El Khoury, C. E. Jabra Alagha, J. A. Skaf and M. A. AlAlaoui, Arabic Speech Recognition Using Recurrent Neural Networks, Faculty of Engineering and Architecture – American University of Beirut
8. Dongsuk YUK, Robust Speech Recognition Using Neural Networks and Hidden Markov Models, The State University of New Jersey
9. Bülent Bolat, Ünal Küçük, Speech Music Classification By Using Statistical Neural Networks, Yıldız Technical University, Department of Electric and Electronic Engineering
10. Paolo Marroe, A Complate Guide All You Need to Know Aboat Joone, 17.1.2007
11. Harshavardhana , Varun Ramesh , Sanjana Sundaresh, Vyshak B N, Speaker Recognition System Using MFCC, A Project Work Of 6th Semester Electronics &Communication Engineering, Visvesvaraya Technological University
12. http://cmusphinx.sourceforge.net/sphinx4/#what_is_sphinx4
13. Audio: A Feature Extraction Library, Daniel McEnnis, Cory McKay, Ichiro Fujinaga, Philippe Depalle, Faculty of Music, McGill University