**International Journal of Recent Development in Engineering and Technology**
**Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 2, Issue 3, March 2014)**

# An Enhanced Speech Recognition System

Suma Shankaranand[1], Manasa S[2], Mani Sharma[3], Nithya A.S[4], Roopa K.S.[5], K.V. Ramakrishnan[6]

*[1]Department of Electronics and Communication Engineering, Research Scholar, Anna University, Chennai,India*
*[2,3,4,5]Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, India*
*[6]Director, AV systems and circuits, Bangalore, India*

***Abstract*— This paper describes the development of an efficient speech recognition system using various techniques such as Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), Hidden Markov Model (HMM) and Autocorrelation.**

**In this paper, a method to recognize the speech faster with more accuracy, speaker recognition is followed by speech recognition. MFCC/Autocorrelation is used to extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Then HMM is used on Quantized feature vectors to identify the word by evaluating the maximum log likelihood values for the spoken word.**

***Keywords*— MFCC, VQ, HMM, log likelihood, Autocorrelation.**

## I. Introduction

Speech recognition research has its roots in the idea of communicating with a machine by voice. Automatic speech recognition uses digital technology for converting speech signals into a sequence of words or other linguistic units by means of an algorithm implemented as a computer program.

Modern speech understanding systems are now capable of understanding speech input for vocabularies of the order of thousands of words in operational environments. Speech signal conveys two important types of information, primarily the speech content and on the secondary level, the speaker identity. Speech recognizers aim to extract the lexical information from the speech signal independently of the speaker by reducing the inter-speaker variability. On the other hand, speaker recognition is concerned with establishing the identity of the person. [3]

The confidence level of authentication needs to be high for biometric security applications, online transactions, database access services, information services, security control for confidential information areas, remote access to computers etc. Speaker identification allows the use of uttered speech to verify the speaker's identity and control access to secure services.

Speech Recognition offers greater freedom to employ the physically handicapped in manufacturing processes, in medicine and in telephone network applications. Speech recognition history dates back to some four decades and yet it has not been accepted as reliable enough to be considered as a standalone security system. This paper focuses on the implementation of speaker identification and enhancement of speech recognition using Hidden Markov Model (HMM) techniques. [1], [4]

## II. History Of Speech Recognition

Speech Recognition research has been ongoing for more than 80 years. Over that period there have been at least 4 generations of approaches, and a 5th generation is being formulated based on current research themes. To cover the complete history of speech recognition is beyond the scope of this paper. The work carried out by different workers from2001 to till date is presented in the following paragraphs.

By 2001, computer speech recognition had topped out at 80% accuracy and near the end of the decade; the technology's progress seemed to be stalled. Speech recognition technology development began to edge back into the forefront with one major event: the arrival of the "Google Voice Search app for the iPhone". In 2010, Google added "personalized recognition" to Voice Search on Android phones, so that the software could record users' voice searches and produce a more accurate speech model. The company also added Voice Search to its Chrome Browser in mid-2011. And now like Google's Voice Search, Apple's "Siri" relies on cloud-based processing. It draws what it knows about you to generate a contextual reply and it responds to your voice input with personality. [2]

Parallel processing methods using combinations of HMMs and acoustic- phonetic approaches to detect and correct linguistic irregularities are used to increase recognition decision reliability and increase robustness for recognition of speech in noisy environment.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 2, Issue 3, March 2014)

## III. Proposed Model

The structure of proposed system consists of two modules namely, Speaker Identification followed by Speech Recognition.

### *A. Speaker Identification*

Feature extraction is a process that extracts data from the voice signal that is unique for each speaker. Mel Frequency Cepstral Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. [6], [7], [8]

These extracted features are Vector quantized using Vector Quantization algorithm. Vector Quantization (VQ) is used for feature extraction in both the training and testing phases. It is an extremely efficient representation of spectral information in the speech signal by mapping the vectors from large vector space to a finite number of regions in the space called clusters. [6], [8]

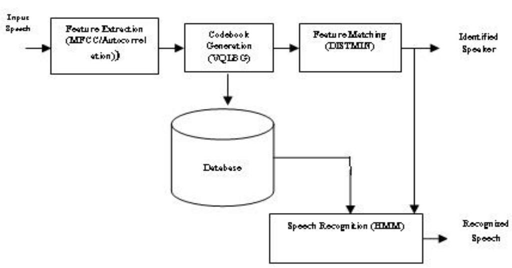


**Figure. 1 Proposed Model**

After feature extraction, feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features with the database using the DISTMIN algorithm.

### *B. Speech Recognition*

Hidden Markov Processes are the statistical models in which one tries to characterize the statistical properties of the signal with the underlying assumption that a signal can be characterized as a random parametric signal of which the parameters can be estimated in a précised and well-defined manner.

In order to implement an isolated word recognition system using HMM, the following steps must be taken

(1) For each uttered word, a Markov model must be built using parameters that optimize the observations of the word.

(2) Maximum likelihood model is calculated for the uttered word. [5], [9], [10], [11]

## IV. Simulation Experiment

A simulation experiment was conducted using Mathworks Matlab Version 7.8.0.347 R2009a software. A database of four different lists of phonetically balanced words as reported in [12] was created for eight male and eight female voices. These lists were randomized for testing phase. Experimentation was performed on the system using the above data. The major modules used are MFCC (Mel-scaled Frequency Cepstral Coefficients), which uses Mel-spaced Filter Bank, VQ (Vector Quantization), HMM (Hidden Markov Model) which uses Discrete - HMM Observation matrix and Forward - Backward algorithm.

The figures 2 and 3 show the GUI (Graphical User Interface) developed to add the voices to the database, test for speaker and speech recognition. The MFCC was then replaced by autocorrelation technique for feature extraction. The speech recognition was compared for both the systems with respect to speed and memory space.

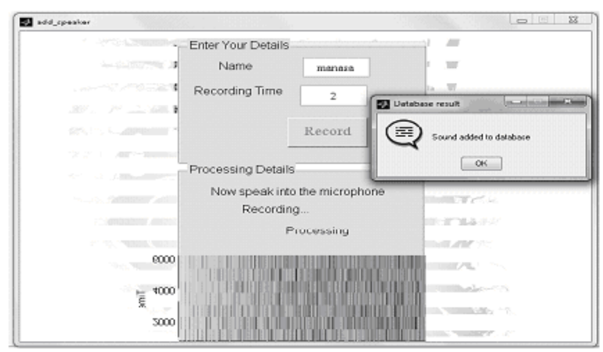


**Figure. 2 Adding a Speaker sound to the database**

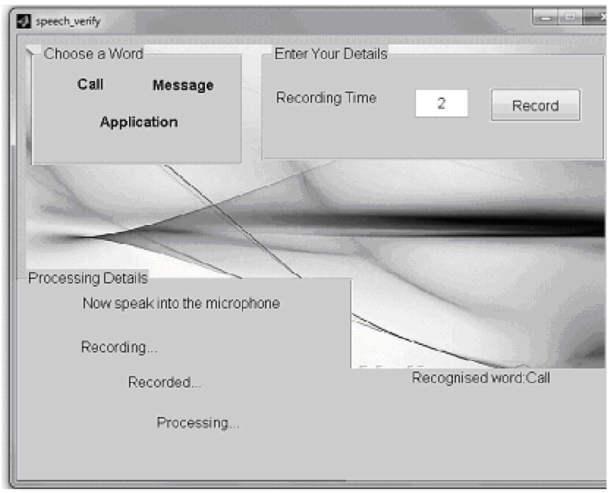


**Figure. 3 Testing the speech recognition system**

## V. EXPERIMENTATION RESULTS

In the speaker identification phase, subject is asked to speak the same word ten times from the given list of words. Sixteen different subjects repeat the same procedure. The subjects are then asked to utter the same set of words in a random order. Time required for speech recognition using MFCC and Autocorrelation were computed. Table 1 shows the computation time for 16 speakers.

**Table 1**
**Comparison Of Computation Time Required For Speech Recognition Using MFCC And Autocorrelation Experiments**

| Sl. No. | Speakers | Time required (in seconds) with MFCC for feature extraction | Time required (in seconds) with Autocorrelation for feature extraction |
|---|---|---|---|
| 1. | Male 1 | 360 | 0.5 |
| 2. | Male 2 | 300 | 0.4 |
| 3. | Male 3 | 320 | 0.5 |
| 4. | Male 4 | 300 | 0.5 |
| 5. | Male 5 | 320 | 0.4 |
| 6. | Male 6 | 360 | 0.5 |
| 7. | Male 7 | 340 | 0.5 |
| 8. | Male 8 | 300 | 0.4 |
| 9. | Female 1 | 300 | 0.5 |
| 10. | Female 2 | 240 | 0.4 |
| 11. | Female 3 | 240 | 0.5 |
| 12. | Female 4 | 260 | 0.4 |
| 13. | Female 5 | 300 | 0.4 |
| 14. | Female 6 | 240 | 0.5 |
| 15. | Female 7 | 300 | 0.4 |
| 16. | Female 8 | 240 | 0.4 |

## VI. CONCLUSION

In the identification of speaker phase, MFCC and Distance Minimum techniques have been used. These two techniques provided more efficient speaker identification system. The speech recognition phase uses the most efficient HMM Algorithm. Speaker recognition module improves the efficiency of speech recognition scores. The coding of all the techniques mentioned above has been done using MATLAB. It has been found that the combination of MFCC and Distance Minimum algorithm gives the best performance and also accurate results in most of the cases with the efficiency of 95%. The study also reveals that the HMM algorithm is able to identify the phonetically balanced isolated word. Speech recognition system achieves 98% efficiency. Autocorrelation technique is best suited for noisy environments. It also takes less memory space and time for computation.

*Acknowledgement*

We acknowledge Visvesvaraya Technological University, Belgaum and Anna University, Chennai for the encouragement and permission to publish this paper. We would like to thank the Principal of Sir MVIT, Dr. M. S. Indira for her support. Our special thanks to Prof. Dilip K. Sen, HOD, department of CSE for his valuable suggestions from time to time.

## REFERENCES

[1] Ronald, M. Baecker, "Readings in human-computer interaction: toward the year 2000", 1995.

[2] Melanie, Pinola, "Speech Recognition Through the Decades: How We Ended Up With Siri", PCWorld.

[3] Ganesh Tiwari, "Text Prompted Remote Speaker Authentication: Joint Speech and Speaker Recognition/Verification System".

[4] Dr.Ravi Sankar, Tanmoy Islam, Srikanth Mangayyagari, "Robust Speech/Speaker Recognition Systems".

[5] Bassam A.Q.Al-Qatab and Raja.N.Aninon, "Arabic Speech Recognition using Hidden Markov Model ToolKit (HTK)", IEEE Information Technology (ITSim), 2010, pp. 557-562.

[6] Ahsanul Kabir, Sheikh Mohammad Masudul Ahsan, "Vector Quantization in Text Dependent Automatic Speaker Recognition using Mel-Frequency Cepstrum Coefficient", 6th WSEAS International Conference on circuits, systems, electronics, control & signal processing, Cairo, Egypt, dec 29-31, 2007, pp. 352-355

[7] Lindasalwa Muda, Mumtaj Begam and Elamvazuthi.,"Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and DTW Techniques ",Journal of Computing, Volume 2, Issue 3, March 2010

[8] Mahdi Shaneh and Azizollah Taheri, "Voice Command Recognition System based on MFCC and VQ Algorithms", World Academy of Science, Engineering and Technology Journal, 2009.

[9] Remzi, Serdar Kurcan, "Isolated word recognition from in-ear microphone data using hidden Markov models (hmm)", Master's Thesis, 2006.

[10] Nikolai Shokhirev, "Hidden Markov Models ", 2010.

[11] L.R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in Speech Recognition", Proceedings of the IEEE Journal, Feb 1989, vol. 77, Issue: 2.

[12] Four phonetically balanced words list, Available from: http://www.meyersound.com/support/papers/speech/pblist.htm