AN EFFICIENT SPEECH RECOGNITION SYSTEM

Suma Swamy[1] and K.V Ramakrishnan

[1]Department of Electronics and Communication Engineering, Research Scholar, Anna University, Chennai
suma_swamy@yahoo.com, ramradhain@yahoo.com

ABSTRACT

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

This paper explains how speaker recognition followed by speech recognition is used to recognize the speech faster, efficiently and accurately. MFCC 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, DISTMIN.

1. INTRODUCTION

The idea of human machine interaction led to research in Speech recognition. Automatic speech recognition uses the process and related technology for converting speech signals into a sequence of words or other linguistic units by means of an algorithm implemented as a computer program. Speech understanding systems presently are capable of understanding speech input for vocabularies of thousands of words in operational environments. Speech signal conveys two important types of information: (a) speech content and (b) The speaker identity. Speech recognisers aim to extract the lexical information from the speech signal independently of the speaker by reducing the inter-speaker variability. Speaker recognition is concerned with extracting the identity of the person. [3]

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 several applications like manufacturing processes, medicine and telephone network. Figure 1(a) shows the speech recognition system without speaker identification. Figure 1(b) shows how the speaker identification followed by speech recognition improves the efficiency. With this approach, the database will be divided into smaller divisions (SP1 to SPn) with respect to different speakers. Hence the speech recognition rate improves for the corresponding speaker.

DOI : 10.5121/cseij.2013.3403

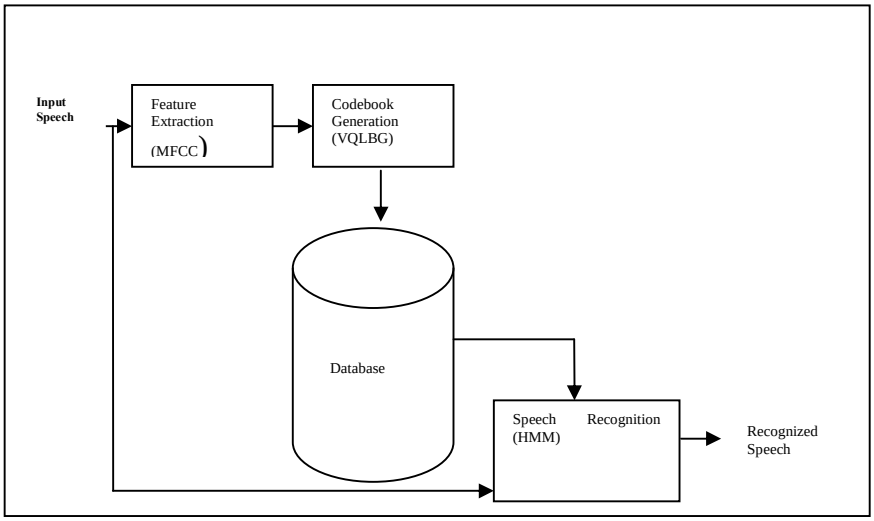


Figure.1 (a) Speech recognition system without speaker identification

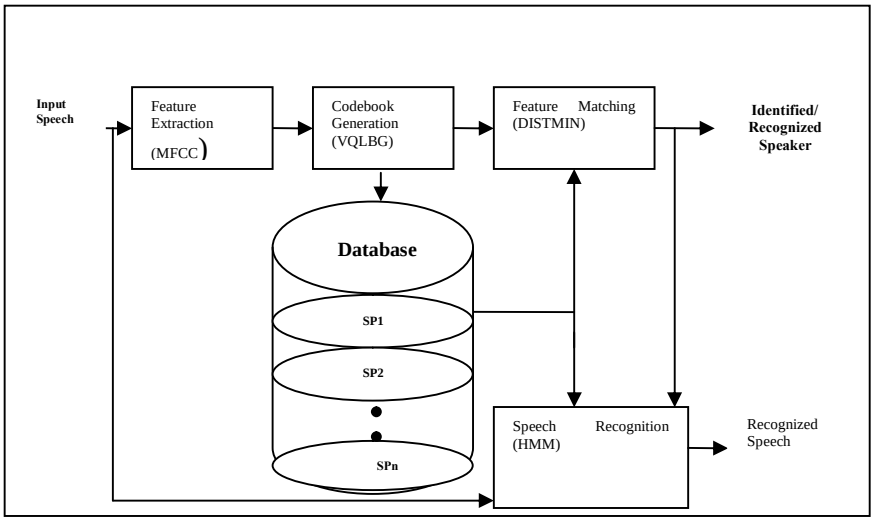


Figure.1 (b) Speaker identification followed by speech recognition

This paper focuses on the implementation of speaker identification and enhancement of speech recognition using Hidden Markov Model (HMM) techniques. [1], [4]

2. 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.

By 2001, computer speech recognition had reached 80% accuracy and no further progress was reported till 2010. 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. Like Google's Voice Search, Siri relies on cloud-based processing. It draws on its knowledge about the speaker to generate a contextual reply and responds to voice input. [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.

3. PROPOSED MODEL

The structure of proposed system consists of two modules as shown in figure 1(b).

- Speaker Identification
- Speech Recognition

3.1 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]
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.

3.2 Speech Recognition System

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 precise 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]

4. IMPLEMENTATION

The major modules used are as follows:

MFCC (Mel-scaled Frequency Cepstral Coefficients)

- Mel-spaced Filter Bank

VQ (Vector Quantization)

HMM (Hidden Markov Model)

- Discrete-HMM Observation matrix
- Forward-Backward Algorithm

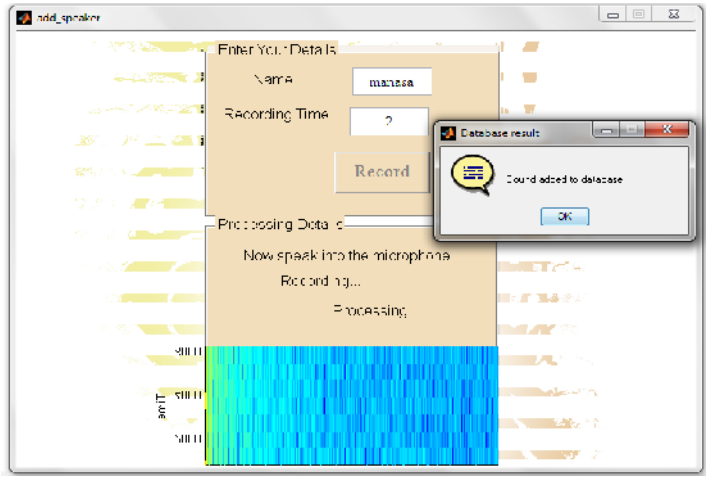


Figure. 2 new speakers sound is added to database by entering name and recording time

Figure 2 shows the screenshot of how new speaker is added into the database. Figure 3 shows the screenshot of how the speech is recognized.

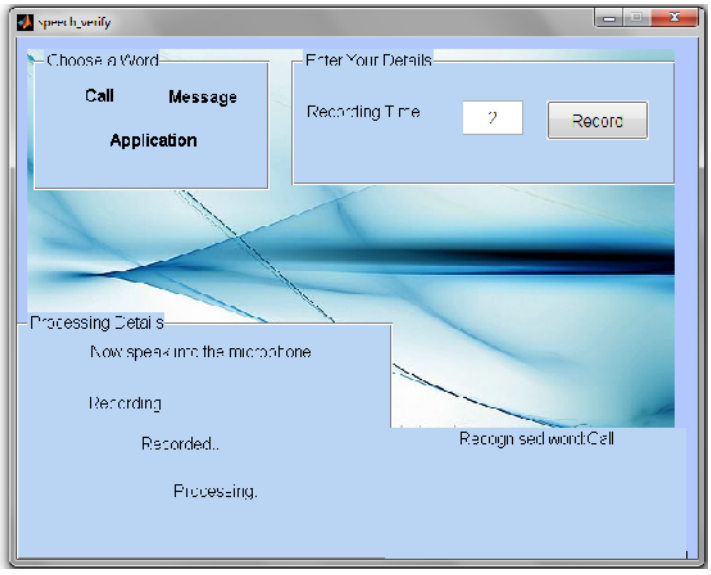


Figure. 3 Screen shot shows the demonstration of speech recognition system

5. EXPERIMENTATION RESULTS

In the speaker identification phase, four different (2 male and 2 female) speakers are asked to speak the same word ten times from the given list of words. The speakers are then asked to utter the same words in a random order and the recognition results noted. The percentage recognition of a speaker for these words is given in the table 1 and efficiency chart is shown in figure 5 for the same. The overall efficiency of speaker identification system is 95%.

In speech recognition phase, the experiment is repeated ten times for each of the above words. The resulting efficiency percentage and its corresponding efficiency chart are shown in table 2 and figure 6 respectively. The overall efficiency of a speech recognition system obtained is 98%.

Table 1. Speaker identification results

Words	Female Speaker 1	Female Speaker 2	Male Speaker 3	Male Speaker 4
Computer	90%	100%	100%	90%
Read	100%	100%	100%	100%
Mobile	90%	100%	90%	90%
Man	100%	70%	100%	100%
Robo	80%	100%	100%	100%
Average %	**92%**	**94%**	**98%**	**96%**

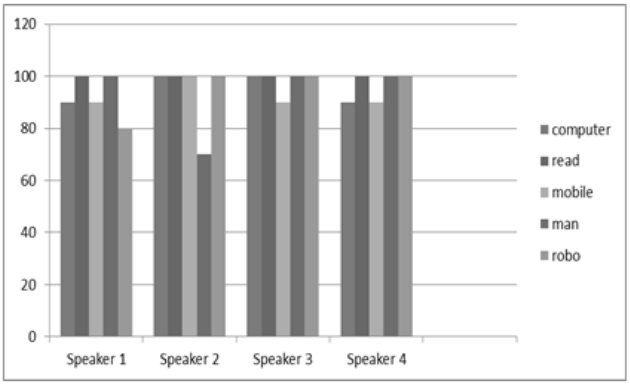


Figure. 4 Efficiency chart for Speaker Identification System

Table 2. Speech Recognition Results

Words	Recognition %
Computer	99%
Read	100%
Mobile	96%
Man	100%
Robo	95%
Average %	**98%**

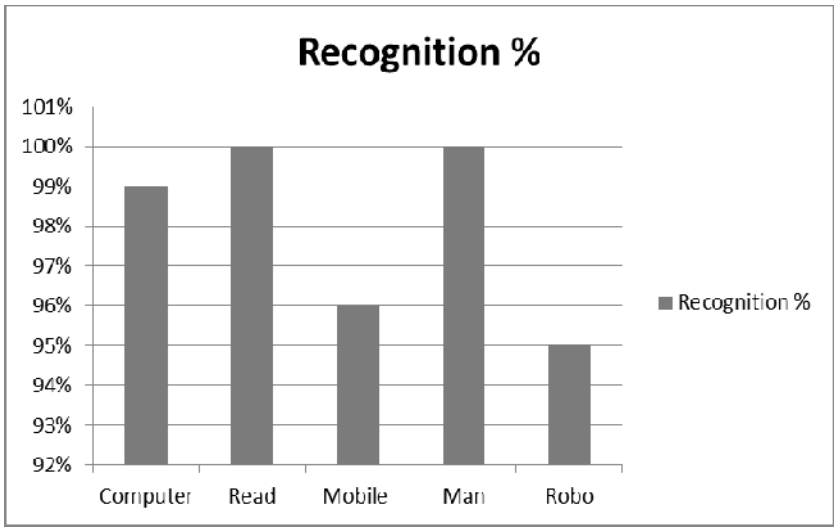


Figure. 5. Efficiency Chart for Speech Recognition System

6. CONCLUSION

In the speaker identification 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. It is found that 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 an overall efficiency of 95%. The study also reveals that the HMM algorithm is able to identify the most commonly used isolated word. As a result of this, speech recognition system achieves 98% efficiency.

ACKNOWLEDGEMENTS

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 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,page 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,page 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.

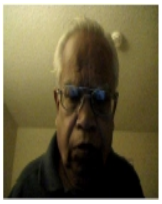
[10] Nikolai Shokhirev ,"Hidden Markov Models ", 2010.

[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] Suma Swamy, Manasa S, Mani Sharma, Nithya A.S, Roopa K.S and K.V Ramakrishnan, "An Improved Speech Recognition System", LNICST Springer Journal, 2013.

AUTHORS

1. Suma Swamy obtained her B.E (Electronics Engineering) in 1990 from Shivaji University, Kolhapur, Maharashtra, and M.Tech (Electronics and Communication Engineering) in 2005 from Visvesvaraya Technological University, Belgaum, Karnataka. She is working as Associate Professor, Department of CSE, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India. She is Research Scholar in the department of ECE, Anna University, Chennai, India. Her areas of interest are Speech Recognition, Database Management Systems and Design of Algorithms.



2. Dr. K.V Ramakrishnan obtained his M.Sc. (Electronics) from Poona University in 1961 and Ph.D (Electronics) from Toulouse (France) in 1972. He worked as Scientist in CEERI from 1962-1999 at different places. He was a Consultant for M/s. Servo Electronics, Delhi in 1999. He was a Director for Research and Development, HOD (ECE/TE/MCA) at Sir M. Visvesvaraya Institute of Technology, Bengaluru, India from 1999 -2002. He was a HOD (ECE) at New Horizon College of Engineering, Bangalore from 2002-03. He was a HOD(CSE/ISE) at Sir M. Visvesvaraya Institute of Technology, Bengaluru, India from 2003- 2006. He was a Dean and Professor (E CE) at CMR Institute of Technology Bangalore from 2006-2009. He also officiated as principal during 2007. He is a Supervisor, Anna University Chennai, India. His area of research is Speech Processing and Embedded Systems.