Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Engineering

Elixir Comp. Engg. 85 (2015) 34194-34200

ISSN: 2229-712X

Performance of Different Configurations of Speech Recognition System

Suma Swamy[1], Anju S Raj[1], H R Ramya[1] and K V Ramakrishnan[2]

[1]Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, India.

[2]AV Systems and Circuits, Bengaluru, India.

ARTICLE INFO

Article history:

Received: 12 June 2015;

Received in revised form:

27 July 2015;

Accepted: 3 August 2015;

Keywords

VQ,
HMM,
MFCC,
LPC,
ANN,
Autocorrelation.

ABSTRACT

This paper discusses in detail different configurations for comparing the performances of different combinations of commonly used techniques for speech recognition and speaker identification. The Feature extraction techniques considered were MFCC, LPC and Autocorrelation. The feature matching techniques considered were VQ, HMM, DTW and ANN.

Introduction

Speech Recognition is a multileveled pattern recognition task. The acoustical signals are examined and structured into a hierarchy of sub-word units e.g., phonemes, words, phrases, and sentences. Each level provides additional temporal constraints, e.g., known word pronunciations or legal word sequences. This can compensate for errors or uncertainties at lower levels. Decisions are combined probabilistically at all lower levels. Discrete decisions are made only at the highest level that can best exploit this hierarchy of constraints. [13]

Speaker Recognition

Speaker recognition is a method of automatically identifying the speaker on the basis of individual information integrated in speech sounds. Speaker recognition finds wide applications to verify the speaker identity and control access to different services viz., banking by telephone, database access services etc.

An important application of speaker recognition is to create new services that will make our everyday lives more secure. Another important application is in forensics. This area has attracted a large number of research workers for the past few decades. Still a number of problems remain unsolved. [16]

Basics of Speaker Recognition

Speaker recognition is the task of recognizing people from their voices. Such systems extract features from speech, model them and use them to recognize the person from his/her voice. Speech recognition becomes more efficient if speaker recognition is done prior to it. This reduces the database volume as it will be restricted to a particular speaker which in turn will reduce the time of searching. Once the native of the speaker is identified, according to the region of the speaker, the database corresponding to only that region accent can be searched.

Speaker recognition has a history dating back from four decades. Speaker recognition uses the acoustic features of speech that differ between individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat and mouth) and learned behavioral patterns (e.g., voice pitch, speaking style). This incorporation of learned patterns (voiceprint) into the voice templates has earned speaker recognition its classification as a "behavioral biometric". [3]

Each speaker recognition system has two phases: 1.Enrolment 2.Test. During enrolment the speaker's voice is recorded and a number of features are derived to form a voiceprint, template, or model. In the test phase (verification or identification phase), the speaker's voice is matched to the templates or models. Speaker recognition systems employ three styles of spoken input: text-dependent, text-prompted and text-independent. This compares spoken text used during enrolment versus test. [10]

Speech Recognition

Speech Recognition System

The block diagram of speech recognition system is shown in Figure 1. The pre-processing involves normalization, parameterization and feature extraction. Different methods used for feature extraction are MFCC, LPC and autocorrelation [6].

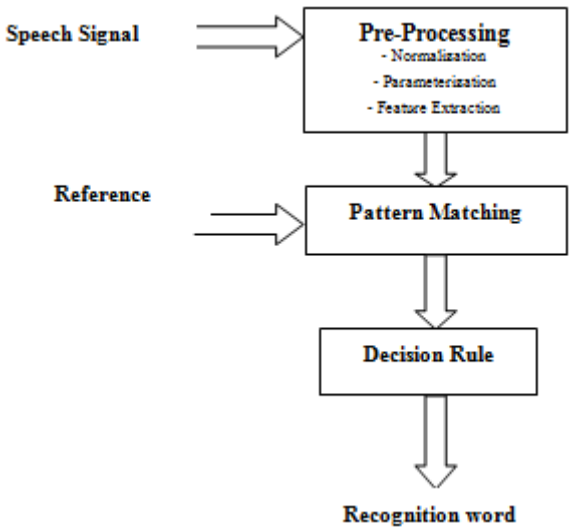

Figure 1. Block Diagram of Speech Recognition [6]

Tele:
E-mail addresses: sumaswamy10@gmail.com

Speech recognition can be classified into text dependent, text-prompted and text independent. In text dependent speech recognition the speaker is prompted to read a given text while in text independent recognition the speaker is free to speak anything he/she wants, clearly.

Text Dependent Speech Recognition

In text dependent recognition, the text is same for enrolment and test. In this mode of verification, the user is expected to say a pre-determined text. It is based on the speaker characteristics as well as the lexical content of the text. As a result, it is more robust and achieves good performance. [4]

The highest accuracies can be achieved if the text to be spoken is fixed. This has the advantage that the system designer can devise a text, which emphasizes speaker differences. However, since the text is always the same such systems are vulnerable to attack by impostors. Furthermore, it is not very user friendly if all users have to remember the same set of text. In addition it makes the system language dependent. But the advantage is that once the region of the speaker is identified, the database search with respect to the accent of the native becomes easier.

Another type of text-dependent system uses pass phrases. The user is free to pick a phrase during enrolment but must use the same phrase during test. Most speaker verification applications use this type of text dependent input. It has the advantage that an impostor must know the pass phrase, which adds a level of security. However, such systems are still vulnerable to tape recorder attacks. [3]

Text Prompted Speech Recognition

In text-prompted systems, the speaker is prompted to utter a given text. This complicates the recognition process. The system must know the text given and must know how this randomly selected text would sound, if spoken by a particular speaker. [3]

Text Independent Speech Recognition

In text independent mode, the system relies only on the voice characteristics of the speaker. The lexical content of the utterance is not used. System models the characteristics of the speaker irrespective of, what one is saying. [4]

Text-independent systems are most often used for speaker identification, as they require very little if any, cooperation by the speaker. In this case the text during enrolment and test is different. [3]

Components of Speech Recognition

Most computer systems for speech recognition include the following five components:

1. Speech Capture Device

This usually consists of a microphone and associated analog-to-digital converter, which digitally encodes the raw speech waveform.

2. Digital Signal Processing Module

The DSP module performs endpoint (word boundary) detection to separate speech from no speech, converts the raw waveform into a frequency domain representation and performs further windowing, scaling, filtering and data compression. The goal is to enhance and retain only those components of the spectral representation that are useful for recognition purposes thereby reducing the amount of information that the pattern-matching algorithm must contend with.

3. Pre processed Signal Storage

The pre-processed speech is buffered for the recognition algorithm.

4. Reference Speech Patterns

Stored reference patterns can be matched against the user's speech sample once the DSP module has pre-processed it. This information is stored as a set of speech templates.

5. Pattern Matching Algorithm

The algorithm must compute a measure of goodness-of-fit between the pre-processed signal from the user's speech and all the stored templates. A selection process chooses the template with the best match.

Two major types of pattern matching in use are template matching by Dynamic Time Warping (DTW) and Hidden Markov Model (HMM).

Template matching by dynamic time warping became very popular in the 1970s. Template matching is conceptually simple. We compare the pre-processed speech waveform directly against a reference template by summing the distances between respective speech frames.

Hidden Markov models are used in most current research systems as this technique produced better results for continuous speech with moderate-size vocabularies. [18]

Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCC)

Block diagram of MFCC process is shown in Figure 2

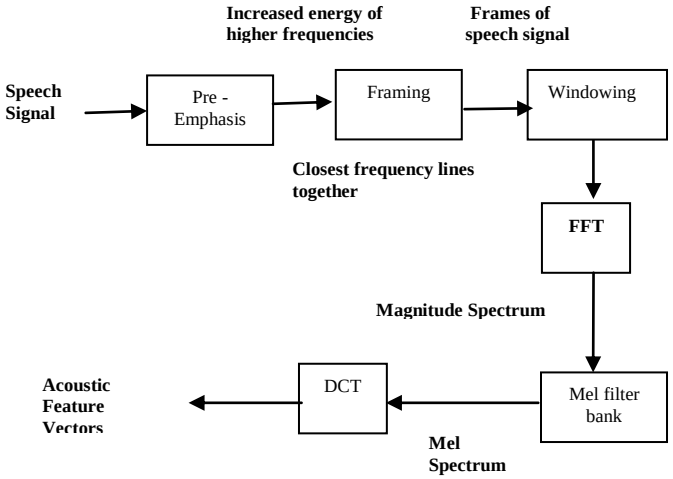

Figure 2. Block diagram of MFCC process[8]

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions i.e known variation of the human ear's critical bandwidth with frequency. MFCC has two types of filters which are spaced linearly at low frequencies below 1000 Hz and logarithmic spacing above 1000 Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetics in speech. It uses the Mel scale expressed on the Mel-frequency scale. The MFCC is the result of a discrete cosine transform of the real logarithm of the short-term energy. [8]

Linear Predictive Coding (LPC)

LPC is based on vocal tract perceptions. It is one of the most powerful speech analysis technique used for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters

LPC is a tool used mostly in speech signal processing for compressed digital signal of speech using the linear predictive model. It is one of the most powerful speech analysis technique used for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters and an efficient computational model of speech. As the actual

predictor coefficients show high variance, they are transformed to a more robust set of parameters known as cepstral coefficients and then used for recognition [see Linear predictive coding]. The LPC model produces parameters that provide a reasonably good representation of the vocal tract. The block diagram of LPC pre-processor stage in speech recognition system is shown in Figure 3.

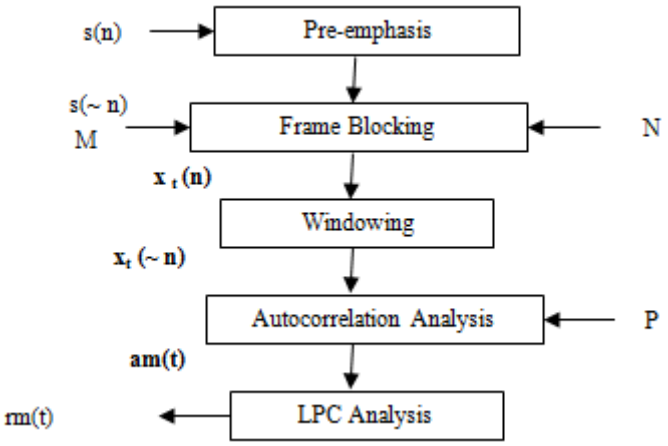

Figure 3. LPC Pre-processor for Speech Recognition [11]

LPC is done in the time domain unlike the other techniques, which are done in the frequency domain. The frequency domain parameters can be easily calculated from the LPC coefficients. [11]

Short Time Autocorrelation Analysis

Autocorrelation refers to the correlation of a time series with its own past and future values. [1]

The autocorrelation function of a signal is basically a (noninvertible) transformation of the signal that is useful for displaying structure in the waveform. Periodicity in the autocorrelation function indicates periodicity in the signal. It is reasonable to define a short-time autocorrelation function, which operates on short segments of the signal. There are two ways of clipping:1.Center clipping2.Infinite peak clipping. Center clipping is used when we need to find the pitch of the signal accurately whereas infinite clipping is used to find the spectral features. [7]

Pattern Matching

Different methods used for pattern matching are Vector Quantization (VQ), artificial neural network- back propagation (BNN), Dynamic Time warping (DTW) and Hidden Markov Model (HMM).

Vector Quantization

VQ is a classical quantization technique, which allows the modeling of probability density functions by the distribution of prototype vectors. It is used for data compression. It works by dividing a large set of vectors into groups having approximately the same number of points closest to them. Each group is represented by its centroid points, as in k-means and other clustering algorithms [20]. A VQ is an "approximator" i.e., it is similar to "rounding off" to the nearest integer. There are various dimensions of VQ like 1-D VQ, 2-D VQ and 3-D VQs. [21]

The goal of pattern recognition is to classify objects of interest into one of a number of classes. The objects of interest are called acoustic vectors that are extracted from an input speech. The classes here refer to individual speakers. Since the classification procedure is applied on extracted features it can also be referred to as feature matching. VQ is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its center called a code word. The collection of all code words is called a codebook. [19]

Designing a codebook that best represents the set of input vectors in NP-hard requires an exhaustive search for the best possible code words in space, and the search increases exponentially as the number of code word increases. We therefore resort to suboptimal codebook design schemes, which are named as LBG (Linde-Buzo-Gray) algorithm. [22]

Back Propagation under Artificial Neural Networks (BNN)

Multi-Layered Perceptron (MLP) has been adapted for speech recognition. A general neural structure is shown in Figure 4. MLP consists of three layers: an input layer, an output layer, and an intermediate or hidden layer. [15]

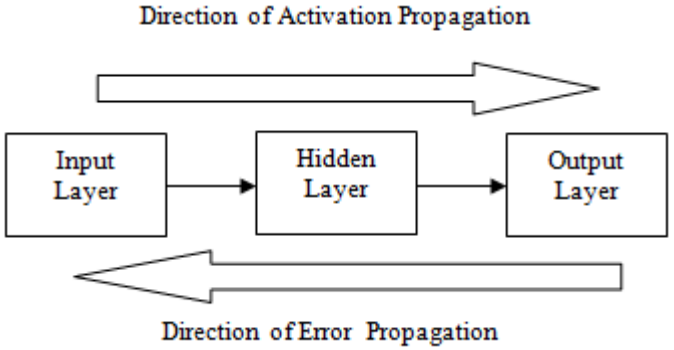

Figure 4. Multilayer Perceptron

Back Propagation Using Hyperbolic Tangent Function

1) Training Set- A collection of input-output patterns is used to train the network.

2) Testing Set-A collection of input-output patterns is used to assess network performance.

3) Learning Rate- A scalar parameter, analogous to step size in numerical integration, is used to set the rate of adjustments.

4) Network Error-Total-Sum-Squared-Error (TSSE) is computed [15]

The Back Propagation algorithm looks for the minimum of the error function in weight space using the method of gradient descent. The combination of weights that minimizes the error function is considered to be a solution of the learning problem.

Dynamic Time Warping (DTW)

After feature extraction of a spoken word, a frame-wise sequence of feature vectors is obtained. The next step is to compare it with set of stored templates for the current speaker. For this, a popular technique called Dynamic Time Warping (DTW) is used. It is a technique, which "warps" the time axis to detect the best match between given two sequences.

Dynamic Time Warping algorithm (DTW) is an algorithm that calculates an optimal warping path between two time series. The algorithm calculates both warping path values between the two series and the distance between them.

Speech is a time-dependent process. Hence the utterances of the same word will have different durations, and utterances of the same word with the same duration will differ in the middle, due to different parts of the words being spoken at different rates. To obtain a global distance between two speech patterns (represented as a sequence of vectors) a time alignment must be performed. [2]

Hidden Markov Model (HMM)

Figure 5 shows hidden Markov model.

The hidden Markov Model is represented by $\lambda = (\pi, A, B)$

where, π = initial state distribution vector, A = State transition probability matrix, B = continuous observation probability density function matrix

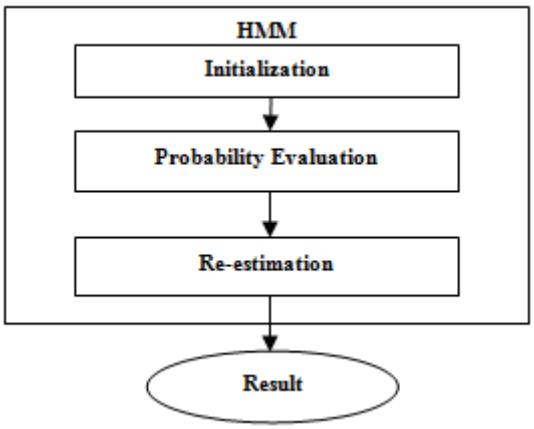

Figure 5. Hidden Markov Model (HMM)

An initial HMM model is used to begin the training process. The initial model can be randomly chosen or selected based on a priori knowledge of the model parameters. The iteration loop is a simple updating procedure for computing the forward and backward model probabilities based on an input speech database (the training set of utterances) and then optimizing the model parameters to give an updated HMM. This process is iterated until no further improvement in probabilities occurs with each new iteration. [5]

Noise Reduction

Different methods used for noise reduction are Voice Activity Detection (VAD), Spectral Subtraction and Cepstral Mean Subtraction (CMS).

Voice Activity Detection (VAD): It is also known as Speech Activity Detection or speech detection. It is a technique used to detect the presence or absence of human speech. This is the first step towards noise reduction. This takes the speech sample as input and returns the sample with the non-speech sections trimmed off. This is done as follows:

1) Speech signal is segmented.
2) The zero-crossing rates for all segments in the speech signal are computed.
3) Frame energy for all segments in the speech signal is computed.
4) Frames with energy greater than ITU (Initial Upper threshold) and frames with energy lesser than ITL (Initial Lower threshold) are searched for.
5) The start and end indices for crossing rates higher than IZCT (Initial Zero Crossing Threshold) are searched for.
6) Thus the speech sample is trimmed using the start and end indices. [12]

Spectral Subtraction is a simple method and effective method of noise reduction. An average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved. [14]

Cepstral Mean Subtraction (CMS) is a normalization method to eliminate channel distortion in speech. It is based on the fact that any convolutional distortion in the time domain transforms to additive distortion in cepstral domain. [17]

Dataset

A database using four different lists consisting of fifty phonetically balanced words as reported in:
http://www.meyersound.com/support/papers/speech/pblist.htm
was created using eight male and eight female voices for testing the different configurations. These words were recorded in a noise free environment. In the training phase, each word was uttered ten times and stored as .wav file. For testing phase the words were randomized. Experiment was performed on all the systems using the above data. All the simulations were done using Mathworks Matlab Version 7.8.0.347 R2009a software.

Different Configurations for Speech Recognition

In order to optimize the performance of speech recognition system, simulation were conducted using different combination of techniques for speaker identification and speech recognition with respect to the aspects like isolated word or continuous speech recognition, speaker independent or speaker dependent, text dependent or text independent, type of vocabulary used, efficiency, speed and memory.

Configuration 1

Figure 6 shows the configuration using MFCC and VQ for both speaker and speech recognition. The system was tested for isolated words and small vocabulary. The efficiency of the system was 95% & 86.67% for speaker and speech recognition respectively. The system was found to be text and speaker dependent.

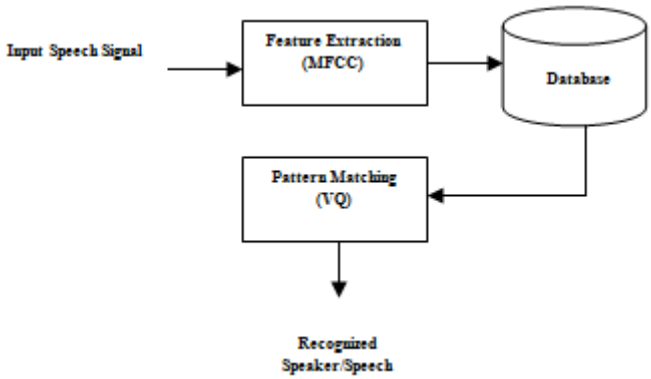

Figure 6. Speaker/speech recognition system using MFCC & VQ

Configuration 2

Figure 7 shows the configuration using MFCC and VQ for speaker identification. LPC and BNN were used for speech recognition. The system gave encouraging results for isolated words and small vocabulary. The system was text and speaker dependent. The efficiency of the system was 95% & 98% for speaker and speech recognition respectively.

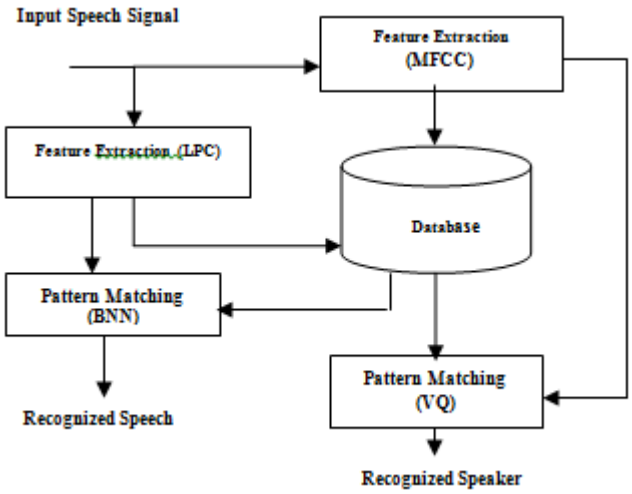

Figure 7. Speaker/Speech Recognition System using MFCC & VQ/ LPC & BNN

Configuration 3

In this configuration, MFCC and DTW were used for speech recognition. Two noise removal algorithms like convolutional noise removal and spectral subtraction were used. Figure 8 shows the configuration. The system gave an efficiency

of 85% for isolated words and small vocabulary and was text dependent and speaker independent.

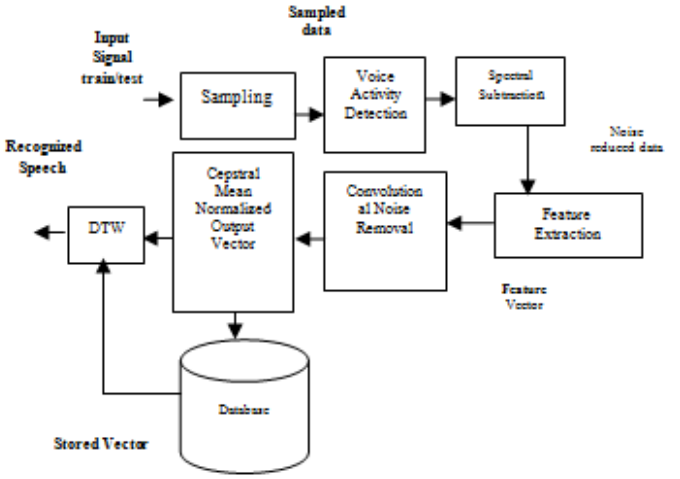

Figure 8. Speech recognition system using MFCC & DTW

Configuration 4

In this configuration, MFCC and VQ were used for speaker recognition. For speech recognition MFCC and HMM were used. The Figure 9 this configuration. The efficiency of the system was 95% & 84.5% for speaker and speech recognition respectively. The system was text and speaker dependent.

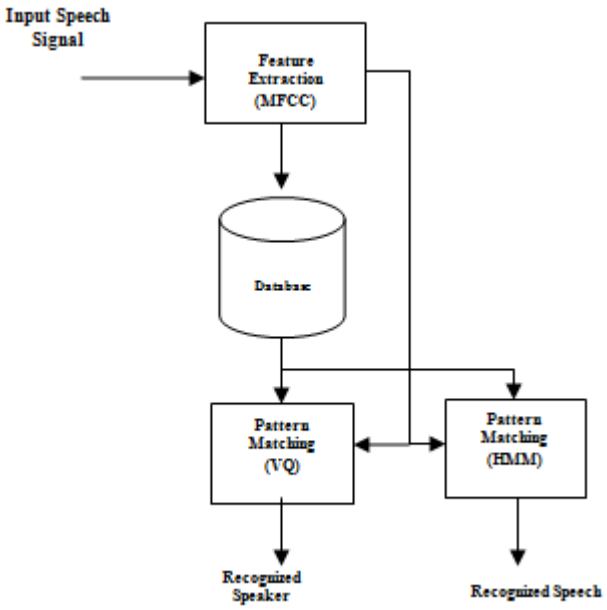

Figure 9: Speaker/Speech recognition system using MFCC, VQ & HMM

Configuration 5

The Figure 10 shows the configuration using MFCC and HMM for speech recognition. End point detection algorithm was used for noise removal. The efficiency of the system was 84.5% for speech recognition but was text dependent and speaker independent.

Parameter values used are: vocabulary size- 200 words, feature dimensions for MFCC-: 16 coefficients, VQ dimensions- 2- dimensional 4 bit ,the number of states in HMM N=3, and the structure of BNN (ANN) - 3 hidden layers.

The table 1 gives the efficiency and application of different configurations. Configuration 1 can be used for isolated word recognition, speaker dependent, text dependent and small vocabulary. The efficiency of this system for speaker and speech recognition is 95% and 86.67% respectively.

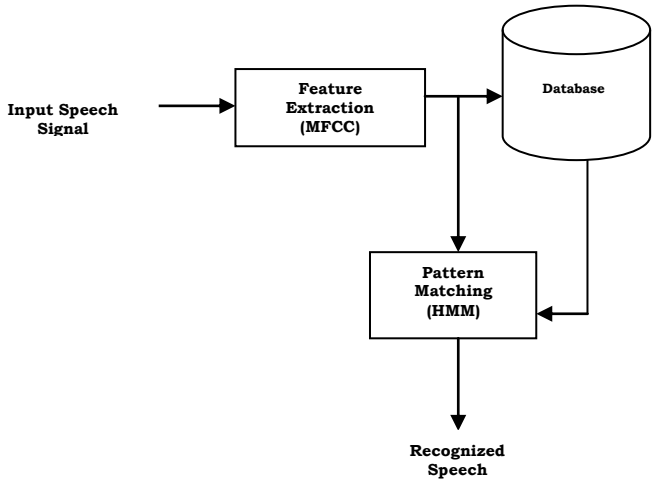

Figure 10. Speech Recognition System using MFCC & HMM

Configuration 2 can be used for isolated word recognition, speaker dependent, text dependent and small vocabulary. The efficiency of this system for speaker and speech recognition is 95% and 98% respectively. Configuration 3 can be used for isolated word recognition, speaker independent, text dependent, noise free and small vocabulary. The efficiency of this system for speech recognition is 85%. Configuration 4 can be used for isolated word recognition, speaker dependent, text dependent and small vocabulary. The efficiency of this system for speaker and speech recognition is 95% and 84.5% respectively. Configuration 5 can be used for isolated word recognition, speaker independent, text dependent, noise free and small vocabulary. The efficiency of this system for speech recognition is 84.5%.

Simulation using autocorrelation

MFCC was replaced by autocorrelation function for feature extraction. Figure 11 shows the block diagram for configuration 5 using autocorrelation function.

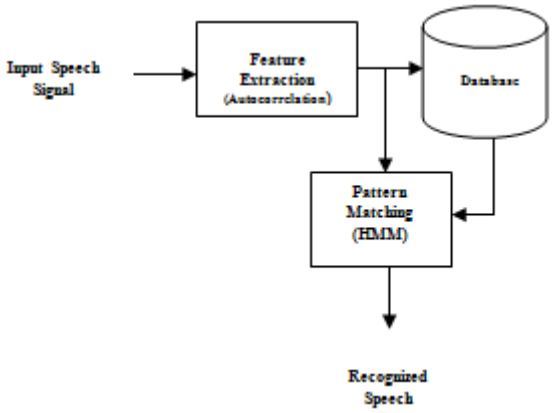

Figure 11. Speech Recognition System with Autocorrelation and HMM

Comparison of MFCC and Autocorrelation

The tests were conducted for both speaker dependent and speaker independent systems. MFCC and autocorrelation functions were used for feature extraction. The time required for recognition of speech were computed and compared. Table 2 indicates the computation time.

It is observed that the time taken by systems using autocorrelation for feature extraction is lower. In case of systems using HMM and VQ it is 500 times faster. In case of GMM and ANN it is 50% faster. To understand why it is so, frequency analysis and time duration were observed for 50 typical words.

Table1. Efficiency and application of different configurations

	Config. 1	Config.2	Config.3	Config4	Config.5
Isolated word /continuous speech	Isolated word	Isolated word	Isolated word	Isolated word	Isolated word
Text Independent/ dependent	Text dependent	Text dependent	Text dependent	Text dependent	Text dependent
Speaker independent/ dependent	Speaker dependent	Speaker dependent	Speaker independent	Speaker dependent	Speaker independent
Vocabulary	Small	Small	Small	Small	Small
Efficiency: Speaker Speech	95% 86.67%	95% 98%	- 85%	95% 84.5%	- 84.5%

Table 2. Comparison of computation time required for speech recognition using MFCC and Autocorrelation configurations

Sl. No	System used for recognition	Time required (in seconds) with MFCC for feature extraction				Time required (in seconds) with Autocorrelation for feature extraction			
		Male1	Male2	Female 1	Female 2	Male1	Male2	Female 1	Female 2
1.	DTW	1320	1200	1440	1400	1	1.5	1	1
2.	HMM	360	300	300	240	0.5	0.4	0.5	0.4
3.	VQ	360	240	240	180	0.5	0.4	0.3	0.2
4.	GMM	12	18	24	30	6	5	12	10
5.	ANN	50	48	54	50	25	22	30	28

Table 3. Speech Signal Time in Seconds

WORDS/ SPEAKER	Female1	Female2	Male1	Male 2
BASK	0.5	0.3	0.5	0.3
CANE	0.4	0.4	0.3	0.3
DEED	0.2	0.3	0.5	0.5
DIKE	0.2	0.3	0.5	0.5
DISH	0.3	0.25	0.3	0.4

Table 3 gives the time duration in seconds of speech signal of some words.

The spectrum and spectrogram of fifty words for eight male and eight female voice was generated. This was done using simulink model. A sample for one male and one female for word "Cane" is shown in Figure 12.

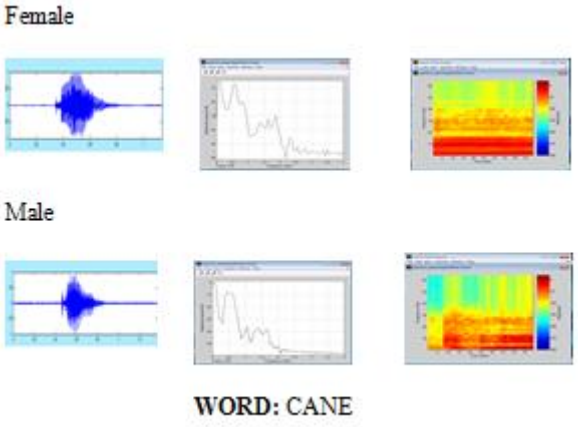

Figure 12. Speech signal, spectrum and spectrogram for "CANE"

The time duration of speech signal for the corresponding word does not vary much compared to spectrum and spectrogram. For the same set of phonetically balanced words uttered by different speakers the spectrum varies quite significantly. This leads to a conclusion that for recognition of speech signal, time analysis is best suited whereas for speaker recognition frequency analysis is better. words, the efficiency of 97% and 93% was achieved for speaker recognition and speech recognition respectively. Autocorrelation function takes care of noise removal from the speech signal thus improving the efficiency. It also reduces the memory requirement leading to faster operation. Autocorrelation techniques are used for detecting signals in noisy environments. It supports the conclusion that this technique with HMM/VQ performs better and is suitable for smart phone applications.

Conclusion

The objective of this work was to develop a speech recognition system, which is faster, requiring minimum storage space and gives better accuracy. It should also work offline. For speaker recognition, many researchers have used MFCC and LPC for feature extraction. For Speech recognition, MFCC followed by VQ /BNN/ GMM/ DTW/ HMM were used.

All these systems require more memory space and hence require more time for execution. This research work aims to develop a speech recognition system for handheld devices used in noisy environments and achieve improved efficiency. It is restricted not only to communication systems, but also other applications like data collection and remote control of appliances by physically handicapped persons.

When autocorrelation was used for feature extraction, the speaker recognition followed by speech recognition was implemented. For the vocabulary of 1000 words, the efficiency of 97% and 93% was achieved for speaker recognition and speech recognition respectively. Autocorrelation function takes care of noise removal from the speech signal thus improving the efficiency. It also reduces the memory requirement leading to faster operation. Autocorrelation techniques are used for detecting signals in noisy environments. It supports the conclusion that this technique with HMM/VQ performs better and is suitable for smart phone applications.

Future Work

Use of speaker recognition followed by speech recognition system reduces the time to search in the database. As the simulation results obtained are favorable in terms of memory space and speed, it can be practically implemented on any of the

handheld device by developing an application on the standard platforms. Memory requirement can be estimated when autocorrelation function is used for feature extraction in case of clipped/unclipped speech signal.

The speech enabled handheld device can work as remote control for any electrically/ electronically operated devices like television, washing machine, fans, lights, robot etc. without requirement of Internet connection.

Acknowledgment

We acknowledge Visvesvaraya Technological University, Belgavi for the encouragement and permission to publish this paper. Suma, Anju and Ramya thank the Principal of Sir MVIT, Dr. M.S.Indira for her support. Suma, Anju and Ramya also thank Prof. Dilip K. Sen, HOD of CSE for his valuable suggestions from time to time.

References

[1] Autocorrelation 2005, Notes_3, GEOS 585A, Spring (2005), available from: shadow.eas.gatech.edu/~jean/paleo/Meko_Autocorrelation.pdf

[2] Dynamic Time Warping, available from: www.cnel.ufl.edu/~kkale/dtw.htm

[3] Encyclopedia -Speaker Recognition, available from: http://www.statemaster.com/encyclopedia/Speaker-recognition

[4] Ganesh, Tiwari, Madhav, Pandey and Manoj, Shrestha (2011),**"**Text-Prompted Remote Speaker Authentication**"**, Project Report, Tribhuvan University, Institute of Engineering, Pulchowk Campus, Nepal.

[5] Lawrence, R. Rabiner, and Ronald, W. Schafer, "Introduction to Digital Speech Processing"; Foundations and Trends in Signal Processing, 1(1/2) , 1-194, IEEE Transactions on Signal Processing , 53(5) (2007), 1881-1896.

[6] Le, Chau, Giang: "Application of a back-propagation neural network to isolated-word speech recognition", Master's Thesis, Naval Postgraduate school, Monterey, CA (1993).

[7] Li, Tan and Montri, Karnjanadecha,"Pitch Detection Algorithm Autocorrelation Method and AMDF", available from: pdfsearch.asia/montri-2.htm

[8] Lindasalwa, Muda, Mumtaj, Begam , and Elamvazuthi, I., "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques"; Journal Of Computing, 2, 3 (2010),138-143.

[9] Linear predictive coding, available from: http://en.wikipedia.org/wiki/Linear_predictive_coding

[10] Maia, E., M., "Speaker identification for Hearing Instruments", Master's Thesis, IMM, Denmarks Technical University (2005).

[11] Milan, G., Mehta, "Speech Recognition System", Master Thesis, Texas Tech University (1996).

[12] Nada, A. G. H. Shindala, "Investigation of distance effect on Gaussian Mixture Models in Speaker Identification"; Al-Rafidain Engineering, 19, 5 (2011), 53-65.

[13] "Neuro AI: Speech Recognition (2013)", available from: http://www.learnartificialneuralnetworks.com/speechrecognition.html

[14] "Noise Reduction**"**, available from: http://sound.eti.pg.gda.pl/denoise/noise.html

[15] Pawar, R.V., Kajave, P.P., and Mali, S. N., "Speaker Identification using Neural Networks"; World Academy of Science, Engineering and Technology,1, 12 (2005), 31-35.

[16] Priyanka, Mishra , and Suyash, Agrawal, "Recognition of Speaker Using Mel Frequency Cepstral Coefficient & Vector Quantization for Authentication**"**;IJSER(International Journal of Scientific & Engineering Research), 3,8 (2012),1-6.

[17] Pu, Yang, Yingchun, Yang, and Zhaohui, Wu, "Robust Speaker Recognition Using Glottal Information Based Cepstral Mean Subtraction"; Proceedings of 148th meeting of the Acoustic Society of America, San Diego, California (2004).

[18] Richard, D., Peacocke, and Daryl, H., Graf, "An Introduction to Speech and Speaker Recognition"; IEEE Computer Journal, 23, 8 (1990), 26-33.

[19] Srinivasa, Kumar, Ch., and Mallikarjuna, Rao P.: "Design of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm"; IJCSE (International Journal on Computer Science and Engineering), 2942-2954.

[20] "Vector quantization", available from: http://en.wikipedia.org/wiki/Vector_Quantization

[21] "Vector Quantization", available from: http://www.data-compression.com/vq.html

[22] "Vector Quantization", available from: http://www.mqasem.net/vectorquantization/vq.html