A Survey on English Digit Speech Recognition using HMM

Vaibhavi Trivedi[1]

[1]Gujarat Technological University,
Department of Master of Computer Engineering,
Noble Engineering College,
Parth Vatika, Near Bamangam, Junagadh 362001
trivedi.vaibhavi@gmail.com

Abstract: ***Speech technology and systems in human computer interaction have witnessed a stable and remarkable advancement over the last two decades. Today, speech technologies are commercially available for an unlimited but an interesting range of tasks. These technologies enable machines to respond correctly and reliably to human voices, and give useful and valuable services. Speech recognition system recognizes the speech samples. Recognition phase of Speech Recognition Process using Hidden Markov Model. Preprocessing, Feature Extraction and Recognition three steps and Hidden Markov Model (used in recognition phase) are used to complete Automatic Speech Recognition System. Hidden Markov Model (HMM) provides a highly reliable way for recognizing speech. The system is able to recognize the speech waveform by translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique This paper focuses on all English digits from (Zero through Nine), which is based on isolated words structure.***

Keywords: Speech Recognition, HMM, MFCC, LPC

1. Introduction

Humans interact with their atmosphere in many ways and receive information through many modalities: sight, audio, smell and touch. To communicate with the atmosphere humans send out signals or information visually, auditory and through gestures. With the development of technology the human dependency on the machines has increased manifold. Humans have to interact with the machines to get the data processed. Human–computer interaction often uses a mouse, keyboard …etc as machine input and screen, printer, speaker...etc as output.

Physically challenged people find computer difficult to use. Partially blind people discover reading from monitor difficult. Moreover current computer interface assumes a certain level literacy from the user. It expects the user to have certain level of proficiency in English. Speech interface can help us deal with these problems. Speech synthesis and speech recognition together form a speech interface. Speech synthesizer converts text into speech. Speech recognition in a computer domain may be defined as the ability of computer systems to accept spoken words in an audio format such as wav, raw recognize the audio format and take appropriate action of operation.

Speech is a complicated biometric signal produces as a result of several transformations occurring at semantic, linguistic, acoustic and articulatory level.

The task of recognition is not easy one the variation in dialect, speaking rate, vocal tract length that exist between speakers account for many of the difficulties encountered during recognition.

1.1 Application

Speech recognizer would enable more efficient communication for everybody, but especially for children, analphabets and people with disabilities. A speech recognizer could also be a subsystem in a speech-to-speech translator. Some typical applications of such numeral recognition are voice-recognized passwords, voice repertory dialers, automated call-type recognition, call distribution by voice commands, credit card sales validation, speech to text processing, automated data entry etc.

2. Classification of speech recognition system

Speech recognition systems can be separated in several different classes by describing the type of speech utterance, type of speaker model, type of channel and the type of vocabulary that they have the ability to recognize. Speech recognition is becoming more complex and a challenging task because of this variability in the signal.

2.2.1 Types of speech Utterance

An utterance is the vocalization of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences. The types of speech utterance are:

1) Isolated Words: Isolated word recognizers usually require each word to have quiet on both sides of the sample window. It doesn't mean that it accepts single words, but it requires a single utterance at a time. This is very well for situations where the user is required to give only one word responses or commands, but is very unnatural for multiple word inputs. It is moderately simple and easiest to implement because word boundaries are obvious and the words tend to be clearly pronounced which is the major of this type. The disadvantage of this type is choosing different boundaries affects the results.

2) Connected Words: Connected word systems are similar to isolated words, but allow separate utterances to be 'run-together' with a minimal pause between them.

3) Continuous Speech: Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Basically, it's computer aural test. It includes a great deal of "co articulation", where adjacent words run together without pauses or any other apparent division between words. Continuous speech recognition systems are most difficult to create because they must utilize special methods to determine word boundaries. As vocabulary grows larger, confusability between different word sequences grows.

4) Spontaneous Speech: This type of speech is natural and not prepare. An ASR system with spontaneous speech should be able to handle a variety of natural speech features such as words being run together and even slight hesitate. Spontaneous (unprepared) speech may include mispronunciations, false-starts, and non-words.

2.2.2 Types of Speaker Model

All speakers have their special voices, due to their unique physical body and personality. Speech recognition system is broadly classified into two main categories based on speaker models namely speaker dependent and speaker independent.

1) Speaker dependent models

Speaker dependent systems are designed for a specific speaker. They are generally more accurate for the particular speaker, but much less accurate for other speakers. These systems are usually easier to develop, cheaper and more accurate, but not as flexible as speaker adaptive or speaker independent systems.

2) Speaker independent models

Speaker independent systems are designed for variety of speakers. It recognizes the speech patterns of a large group of people. This system is most difficult to develop, most expensive and offers less accuracy than speaker dependent systems. However, they are more flexible.

2.2.3 Types of Vocabulary

The size of vocabulary of a speech recognition system affects the complexity, processing requirements and the accuracy of the system. Some applications only require a few words (e.g. numbers only), others require very large dictionaries (e.g. dictation machines). In ASR systems the types of vocabularies can be classified as follows.

- Small vocabulary - tens of words
- Medium vocabulary - hundreds of words
- Large vocabulary - thousands of words
- Very-large vocabulary - tens of thousands of words
- Out-of-Vocabulary- Mapping a word from the vocabulary into the unknown word

Apart from the above characteristics, the environment variability, channel variability, speaking style, sex, age, speed of speech also makes the ASR system more complex. But the efficient ASR systems must deal with the variability in the signal.

2.3 Overview of Automatic speech recognition (ASR) system

The task of ASR is to take an acoustic waveform as an input and produce output as a string of words. Basically, the problem of speech recognition can be stated as follows. When given with acoustic observation X = X1,X2…Xn, the goal is to find out the corresponding word sequence W = W1,W2…Wm that has the maximum posterior probability P(W|X) expressed using Bayes theorem as shown in equation (1). The following figure 1 shows the overview of ASR system.

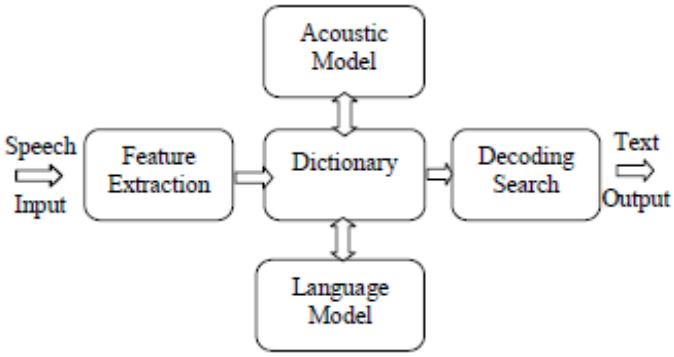

Figure 1: Overview of ASR system

$$W = \arg\max P(W/X) = \arg\max \frac{P(W)P(X/W)}{P(X)}$$

Where P (W) is the probability of word W uttered and P(X|W) is the probability of acoustic observation of X when the word W is uttered.

In order to recognize speech, the system usually consists of two phases. They are called pre-processing and post-processing. Pre-processing involves feature extraction and the post-processing stage comprises of building a speech recognition engine. Speech recognition engine usually consists of knowledge about building an acoustic model, dictionary and grammar. Once all these details are given correctly, the recognition engine identifies the most likely match for the given input, and it returns the recognized word.

An essential task of developing any ASR system is to choose the suitable feature extraction technique and the recognition approach. The suitable feature extraction and recognition technique can produce good accuracy for the given application. Hence, these two major components are reviewed and compared based on its merits and demerits to find out the best technique for speech recognition system. The various types of feature extraction and speech recognition approaches are explained below.

2.4 Speech Feature Extraction Techniques

Feature Extraction is the most important part of speech recognition since it plays an important role to separate one speech from other. Because every speech has different individual characteristics embedded in utterances. These characteristics can be extracted from a wide range of feature extraction techniques proposed and successfully exploited for speech recognition task. But extracted feature should meet some criteria while dealing with the speech signal such as:

- Easy to measure extracted speech features
- It should not be susceptible to mimicry

- It should show little fluctuation from one speaking environment to another
- It should be stable over time
- It should occur frequently and naturally in speech

The most widely used feature extraction techniques are explained below:

Used MFCC And LPC algorithms to extract the features, choose it for the following reasons :

1 One of the most important features, which is required among various kinds of speech applications.

2 Shows high accuracy results for clean speech .

3 They can be regarded as the "standard" features in speaker as well as speech recognition. However, experiments show that the parameterization of the MFC coefficients which is best for discriminating speakers is different from the one usually used for speech recognition applications

3. Mel Frequency Cepstral Coefficient

Inside the MFCC algorithm: This is the block diagram for the feature extraction processes applying mfcc algorithm :

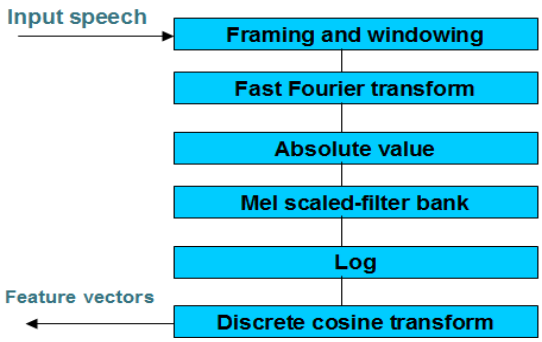

Figure 2: MFCC Flow diagram

- **Preprocessing**

To enhance the accuracy and efficiency of the extraction processes, speech signals are normally pre-processed before features are extracted. Speech signal pre-processing covers digital filtering and speech signal detection. Filtering includes pre-emphasis filter and filtering out any surrounding noise using several algorithms of digital filtering.

Pre-emphasis filter:

In general, the digitized speech waveform has a high dynamic range and suffers from additive noise. In order to reduce this range, pre-emphasis is applied. This pre-emphasis is done by using a first-order FIR high-pass filter.
In the time domain, with input x[n] and $0.9 \leq a \leq 1.0$, the filter equation

$$y[n] = x[n] - a \cdot x[n-1].$$

And the transfer function of the FIR filter in z-domain is:

$$H(Z) = 1 - \alpha \cdot z - 1 , 0.9 \leq \alpha \leq 1.0$$

Where α is the pre-emphasis parameter.

The pre-emphasizer is implemented as a fixed coefficient filter or as an adaptive one, where the coefficient a is adjusted with time according to the auto-correlation values of the speech.

The aim of this stage is to boost the amount of energy in the high frequencies. The drop in energy across frequencies (which is called spectral tilt) is caused by the nature of the glottal pulse. Boosting the high frequency energy makes information from these higher formants available to the acoustic model. The pre-emphasis filter is applied on the input signal before windowing.

- **Framing and windowing:**

first we split the signal up into several frames such that we are analyzing each frame in the short time instead of analyzing the entire signal at once, at the range (10-30) ms the speech signal is for the most part stationary [4].

Also an overlapping is applied to frames. Here we will have something called the Hop Size. In most cases half of the frame size is used for the hop size. The reason for this is because on each individual frame, we will also be applying a hamming window which will get rid of some of the information at the beginning and end of each frame. Overlapping will then reincorporate this information back into our extracted features.

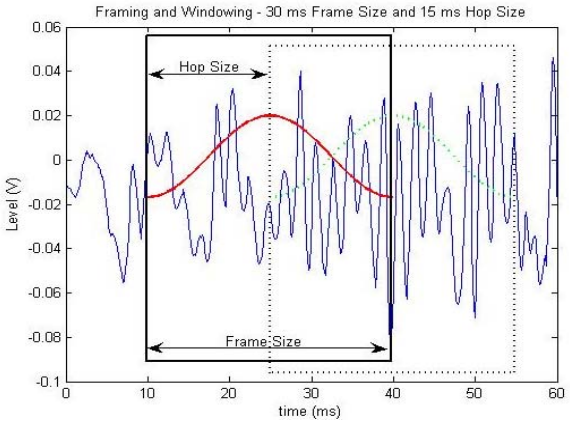

Figure 3: Framing and windowing

- **Windowing**

It is necessary to work with short term or frames of the signal. This is to select a portion of the signal that can reasonably be assumed stationary. Windowing is performed to avoid unnatural discontinuities in the speech segment and distortion in the underlying spectrum [4][5]. The choice of the window is a tradeoff between several factors. In speaker recognition, the most commonly used window shape is the hamming window [6].

The multiplication of the speech wave by the window function has two effects :-

1-It gradually attenuates the amplitude at both ends of extraction interval to prevent an abrupt change at the endpoints

2-It produces the convolution for the Fourier transform of the window function and the speech spectrum.

Actually there are many types of windows such as: Rectangular window, Hamming window, Hann window, Cosine window, Lanczos window, Bartlett window (zero valued end-points), Triangular window (non-zero end-points), Gauss windows …….

We used hamming window the most common one that being used in speaker recognition system.

The hamming window WH (n), defined as [6]:

$$W_H(n) = 0.54 - 0.46\cos\left(\frac{2n\pi}{N-1}\right)$$

The use for hamming windows is due to the fact that mfcc will be used which involves the frequency domain(hamming windows will decrease the possibility of high frequency components in each frame due to such abrupt slicing of the signal).

- **Fast Fourier Transform**

To convert the signal from time domain to frequency domain preparing to the next stage (mel frequency wrapping).
The basis of performing Fourier transform is to convert the convolution of the glottal pulse and the vocal tract impulse response in the time domain into multiplication in the frequency domain [6][7].

Spectral analysis shows that different timbres in speech signals corresponds to different energy distribution over frequencies. Therefore we usually perform FFT to obtain the magnitude frequency response of each frame.

- **Mel-scaled filter bank**

➢ **The mel scale**

The speech signal consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured on the 'Mel' scale. The *mel-frequency* scale is a linear frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. we can use the following formula to compute the mels for a given frequency *f* in Hz:

mel(f)= 2595*log10(1+f/700) [7].

One approach to simulating the subjective spectrum is to use a filter bank, one filter for each desired Mel frequency component. The filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval.

➢ **Mel frequency analysis**

- Mel-Frequency analysis of speech is based on human perception experiments**.**
- Human ears, for frequencies lower than 1 kHz, hears tones with a linear scale instead of logarithmic scale for the frequencies higher than 1 kHz.
- The information carried by low frequency components of the speech signal is more important compared to the high frequency components. In order to place more emphasize on the low frequency components, mel scaling is performed.
- Mel filter banks are non-uniformly spaced on the frequency axis, so we have more filters in the low frequency regions and less number of filters in high frequency regions [2].

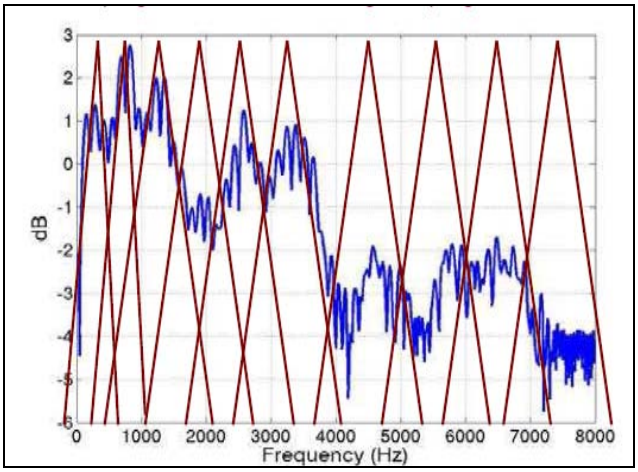

Figure 4: Filter banks

So after having the spectrum (fft for the windowed signal) we applied mel filter banks , the signal processed in such away like that of human ear response:

$$\tilde{S}(l) = \sum_{k=0}^{N/2} S(k) M_l(k)$$

Where :
S(l) :Mel spectrum.
S(K) :Original spectrum.
M(K) :Mel filterbank.
L=0 ,1 , ..., L-1 , Where L is the total number of mel filterbanks
N/2 = Half FFT size.
Now, we will move to the next stage to have the cepstrum or the mel frequency cepstrum coefficients.

➢ **Cepstrum**

In the final step, the log mel spectrum has to be converted back to time. The result is called the mel frequency cepstrum coefficients (MFCCs). The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Because the mel spectrum coefficients are real numbers(and so are their logarithms), they may be converted to the time domain using the Discrete Cosine Transform (DCT).

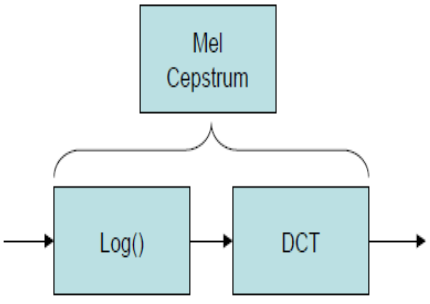

Figure 5: Mel Cepstrum Coefficient

Since the speech signal represented as a convolution between slowly varying vocal tract impulse response (filter) and quickly varying glottal pulse (source), so, the speech spectrum consists of the spectral envelop(low frequency) and the spectral details(high frequency).

Now, our goal is to separate the spectral envelope and spectral details from the spectrum.
It is known that the logarithm has the effect of changing multiplication into addition. Therefore we can simply converts the multiplication of the magnitude of the Fourier transform into addition.

4. Linear Predictive Coding (LPC)

One of the most powerful signal analysis techniques is the method of linear prediction. LPC [5][6] of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it is also an efficient computational model of speech. The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and predicted values, a unique set of parameters or predictor coefficients can be determined. These coefficients form the basis for LPC of speech [7]. The analysis provides the capability for computing the linear prediction model of speech over time. The predictor coefficients are therefore transformed to a more robust set of parameters known as cepstral coefficients. The following figure 6 shows the steps involved in LPC feature extraction.

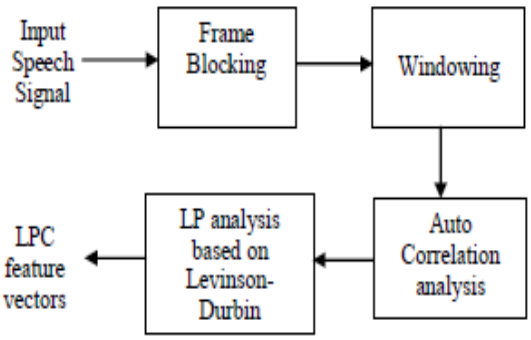

Figure 6: Steps involved in LPC feature extraction

2.5 Speech recognition approaches

In the earlier years, dynamic programming techniques have been developed to solve the pattern-recognition problem [9]. Subsequent researches were based on Artificial Neural Network (ANN) techniques, in which the parallel computing found in biological neural systems is mimicked. More recently, stochastic modeling schemes have been incorporated to solve the speech recognition problem, such as the Hidden Markov Modeling (HMM) approach. At present, much of the recent researches on speech recognition involve recognizing continuous speech from a large vocabulary using HMMs, ANNs, or a hybrid form [9]. These techniques are briefly explained below.

A. Template-Based Approaches

Template based approaches to speech recognition have provided a family of techniques that have advanced the field speech is compared against a set of pre-recorded words (templates) in order to find the best match (Rabiner et al., 1979). This has the advantage of using perfectly accurate word models; but it also has the disadvantage that the pre-recorded templates are fixed, so variations in speech can only be modeled by using many templates per word, which eventually becomes impractical. Template preparation and matching become prohibitively expensive or impractical as vocabulary size increases beyond a few hundred words. This method was rather inefficient in terms of both required storage and processing power needed to perform the matching. Template matching was also heavily speaker dependent and continuous speech recognition was also impossible.

B. Knowledge-Based Approaches

The use of knowledge/rule based approach to speech recognition has been proposed by several researchers and applied to speech recognition (De Mori & Lam, 1986; Alikawa, 1986; Bulot & Nocera, 1989), speech understanding systems (De Mori and Kuhn, 1992). The expert knowledge about variations in speech is hand-coded into a system. It uses set of features from the speech, and then the training system generates set of production rules automatically from the samples. These rules are derived from the parameters that provide most information about a classification. The recognition is performed at the frame level, using an inference engine (Hom, 1991) to execute the decision tree and classify the firing of the rules. This has the advantage of explicitly modeling variations in speech; but unfortunately such expert knowledge is difficult to obtain and use successfully, so this approach was judged to be impractical, and automatic learning procedures were sought instead.

C. Neural Network-Based Approaches

Another approach in acoustic modeling is the use of neural networks. They are capable of solving much more complicated recognition tasks, but do not scale as excellent as Hidden Markov Model (HMM) when it comes to large vocabularies. Rather than being used in general-purpose speech recognition applications they can handle low quality, noisy data and speaker independence [10] [11]. Such systems can achieve greater accuracy than HMM based systems, as long as there is that use the neural network part for phoneme recognition and the HMM part for language modeling.

D. Dynamic Time Warping (DTW)-Based Approaches

Dynamic Time Warping is an algorithm for measuring similarity between two sequences which may vary in time or speed [12]. A well known application has been ASR, to cope with different speaking speeds. In general, it is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions, i.e. the sequences are "warped" non-linearly to match each other. This sequence alignment method is often used in the context of HMM. In general, DTW is a method that allows a computer to find an optimal match between two given sequences (e.g. time series) with certain restrictions. This technique is quite efficient for isolated word recognition and can be modified to recognize connected word also [12].

E. Statistical-Based Approaches

In this approach, variations in speech are modeled statistically (e.g., HMM), using automatic learning procedures. This approach represents the current state of the art. Modern general-purpose speech recognition systems are based on statistical acoustic and language models. Effective acoustic and language models for ASR in unrestricted domain require large amount of acoustic and linguistic data for parameter estimation. Processing of large amounts of training data is a key element in the development of an effective ASR technology nowadays. The main disadvantage of statistical models is that they must make a priori modeling assumptions, which are liable to be inaccurate, handicapping the system's performance.

Hidden Markov Model (HMM)-Based Speech Recognition

The reason why HMMs are popular is because they can be trained automatically and are simple and computationally feasible to use [13] [14]. HMMs to represent complete words can be easily constructed (using the pronunciation dictionary) from phone HMMs and word sequence probabilities added and complete network searched for best path corresponding to the optimal word sequence. HMMs are simple networks that can generate speech (sequences of cepstral vectors) using a number of states for each model and modeling the short-term spectra associated with each state with, usually, mixtures of multivariate Gaussian distributions (the state output distributions). The parameters of the model are the state transition probabilities and the means, variances and mixture weights that characterize the state output distributions. Each word, or each phoneme, will have a different output distribution; a HMM for a sequence of words or phonemes is made by concatenating the individual trained HMM [9] for the separate words and phonemes.

Current HMM-based large vocabulary speech recognition systems are often trained on hundreds of hours of acoustic data. The word sequence and a pronunciation dictionary and the HMM [8] [9] training process can automatically determine word and phone boundary information during training. This means that it is relatively straightforward to use large training corpora. It is the major advantage of HMM which will extremely reduce the time and complexity of recognition process for training large vocabulary.

2.6 Performance Evaluation of ASR Techniques

The performance of a speech recognition system is measurable. Perhaps the most widely used measurement is accuracy and speed. Accuracy is measured with the Word Error Rate (WER), whereas speed is measured with the real time factor. WER can be computed by the equation (1)

$$WER = \frac{S+D+I}{N} \quad (1)$$

Where S is the number of substitutions, D is the number of the deletions, I is the number of the insertions and N is the number of words in the reference.

The speed of a speech recognition system is commonly measured in terms of Real Time Factor (RTF). It takes time *P* to process an input of duration *I*. It is defined by the formula

$$RTF = \frac{P}{I} \quad (2)$$

3.1 Proposed work for Isolated English digit Speech Recognition using HMM

Success of any automatic speech recognition system requires a combination of various and algorithms, each of which performs a specific task for achieving the main goal of the system. Therefore, a combination of related algorithms improves the accuracy or the recognition rate of such applications. the architecture of the HMM based English digits speech recognition system.

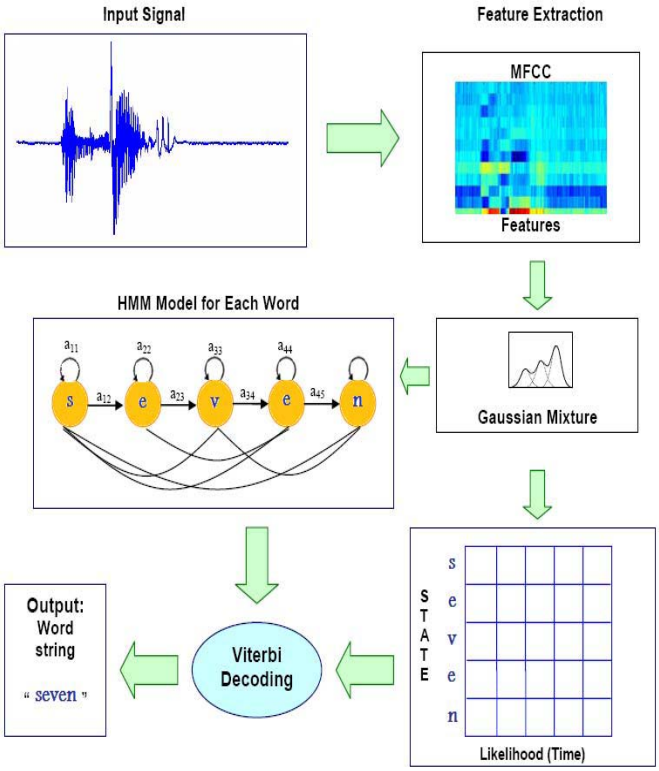

Figure 7: Architecture of HMM based English digit speech recognition

Figure 7 had just shown the main steps to perform the HMM based speech recognition system as follows:

1. Receiving and digitizing the input speech signal.
2. Extracting features for all input speech signals using MFCC algorithm, where its computational steps are shown in Fig. 3 and Fig. 4, then converting and storing each signal's features into a feature vector.
3. Classifying the feature vectors into the phonetic based categories at each frame using HMM algorithm.
4. Finally, performing a Viterbi search which is an algorithm to compute the optimal (most likely) state sequence in HMM given a sequence of observed outputs.

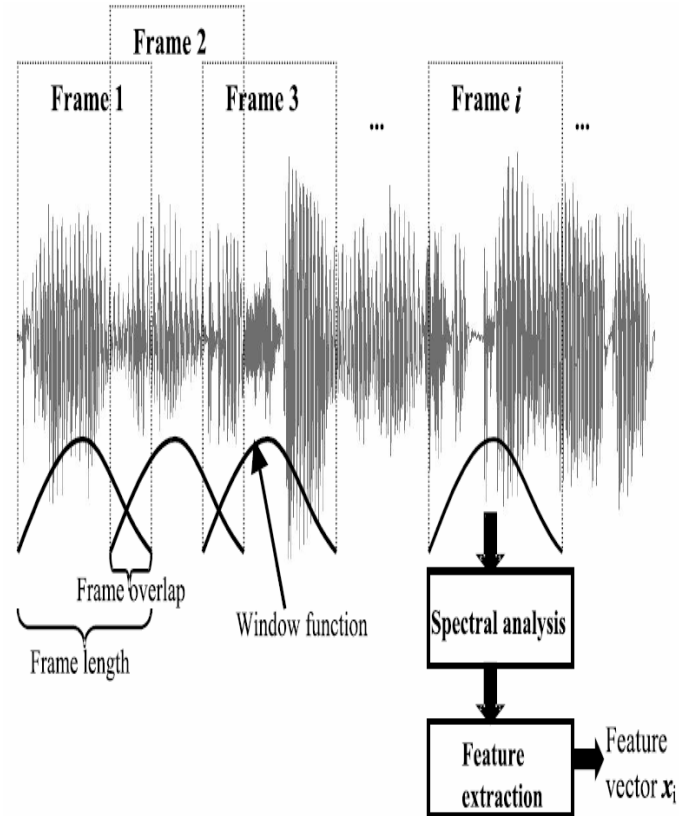

Figure 8: Feature extraction concepts

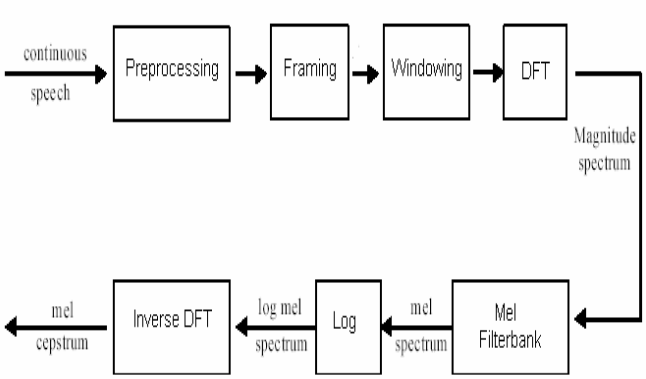

Figure 9: MFCC Computational process

Hidden Markov Model (HMM) is one of the most powerful and dominating statistical approaches, which has been applied for many years. The basic theory of HMM was published in a series of classic papers by Baum and his colleagues in the late 1960s and early 1970s which was then implemented for speech recognition applications by Baker at Carnegie Mellon University (CMU) and by Jelinek and his colleagues at IBM in the 1970s [5]. An HMMs are specified by a set of states *Q*, a set of transition probabilities *A*, a set of observation likelihoods *B*, a defined start state and end state(s), and a set of observation symbols *O*, which is not drawn from the same alphabet as the state set *Q* [6].

References

[1] Bassam A. Q. Al-Qatab , Raja N. Ainon, "Arabic Speech Recognition Using Hidden Markov Model Toolkit(HTK)", 978-1-4244-6716-711 0/$26.00 ©2010 IEEE.

[2] M. Chandrasekar, M. Ponnavaikko, "Tamil speech recognition: a complete model", Electronic Journal «Technical Acoustics» 2008, 20.

[3]M., Forsberg, "Why is Speech Recognition Difficult?". Department of Computing Science, Chalmers University of Technology, Gothenburg, Sweden, 2003.

[4] M. A. M. Abu Shariah, R. N. Ainon, R. Zainuddin, and O. O. Khalifa, "Human Computer Interaction Using Isolated-Words Speech Recognition Technology," IEEE Proceedings of The International Conference on Intelligent and Advanced Systems (ICIAS'07), Kuala Lumpur, Malaysia, pp. 1173 – 1178, 2007.

[5] Corneliu Octavian DUMITRU, Inge GAVAT, "A Comparative Study of Feature Extraction Methods Applied to Continuous Speech Recognition in Romanian Language", 48th International Symposium ELMAR-2006, 07-09 June 2006, Zadar, Croatia.

[6] DOUGLAS O'SHAUGHNESSY, "Interacting With Computers by Voice: Automatic Speech Recognition and Synthesis", Proceedings of the IEEE, VOL. 91, NO. 9, September 2003, 0018-9219/03$17.00 © 2003 IEEE

[7] N.Uma Maheswari, A.P.Kabilan, R.Venkatesh, "A Hybrid model of Neural Network Approach for Speaker independent Word Recognition", International Journal of Computer Theory and Engineering, Vol.2, No.6, December, 2010 1793-8201.

[8] A.P.Henry Charles & G.Devaraj, "Alaigal-A Tamil Speech Recognition", Tamil Internet 2004, Singapore.

[9] Zhao Lishuang , Han Zhiyan, Speech Recognition System Based on Integrating feature and HMM, 2010 International Conference on Measuring Technology and Mechatronics Automation, 978-0-7695- 3962-1/10 $26.00 © 2010 IEEE.

[10] Meysam Mohamad pour, Fardad Farokhi, An Advanced Method for Speech Recognition, World Academy of Science, Engineering and Technology 49 2009

[11] Vimal Krishnan V. R, Athulya Jayakumar and Babu Anto.P, Speech Recognition of Isolated Malayalam Words Using Wavelet Features and Artificial Neural Network, 4th IEEE International Symposium on Electronic Design, Test & Applications, 0-7695-3110-5/08 $25.00 © 2008 IEEE

[12] Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar, A Review on Speech Recognition Technique, International Journal of Computer Applications (0975 8887) Volume 10 No.3, November 2010

[13] M. Chandrasekar, M. Ponnavaikko, Tamil speech recognition: a complete model , Electronic Journal Technical Acoustics 2008, 20

[14] Ghulam Muhammad, Yousef A. Alotaibi, and Mohammad Nurul Huda , Automatic Speech Recognition for Bangia Digits ,Proceedings of 2009 12th International Conference on Computer and Information Technology (ICCIT2009) 21-23 December, 2009, Dhaka, Bangladesh, 978-1-4244-628 1/09/$26.00

Author Profile

Vaibhavi Trivedi received her BE (Computer Engineering in 2007 and M. E. (Computer Engineering-pursing) in 2011-2013.Currently she is researcher student of Noble Engineering college from Gujarat Technological University, Gujarat, India. Her research areas are Speech Recognition and Artificial Intelligence.