A Study of Techniques and Processes Involved in Speech Recognition System

Hrugved Pawar[1], Nikki Gaikwad[2], Shambhavi Kulkarni[3]

[1]Systems Engineer, Infosys Ltd., Phase-3, Hinjewadi IT Park, Pune
[2]Associate, SS&C GlobeOp, Mindspace Airoli, Navi Mumbai
[3]Intern, Algoanalytics Pvt. Ltd., Baner, Pune

---***---

Abstract - *Speech is a simple and usable method of interaction between people, yet these days people are not only restricted to communicating with one another but even to the various machines in our lives. The most significant are the computers and mobiles. This communication is done through interfaces, this area called Human Computer Connection (HCI). Speech recognition is a process to change over speech sound to corresponding text. Speech recognition innovation has been created to an enormous degree in most recent years. Yet at the same time there exist numerous significant examination challenges. For example, speaker and language inconstancy, ecological clamor and the vocabulary size and so forth. The target of this paper is to introduce a total point of view on speech recognition depicting different cycles and summing up different processes utilized in a typical speech recognition system.*

Key Words: Speech recognition, Feature extraction, Modeling, Speech processing, machine control, human machine interaction, MFCC, HMM, Vector quantization.

1. INTRODUCTION

Speech is a characteristic method of communication for individuals, however it is not always effective, for example, in case of debilitated individual. Throughout the most recent multi decade, there has been a need to empower human to speak with machines without performing any text input. Here speech recognition is an innovation that can make it possible for impaired people to speak with machines with their inabilities. Speech recognition is a framework that is utilized by the human to tune in, distinguish and comprehend what the client needs by talking. It is a transformation of speech to message in a system.

Speech recognition is the machine on the statement or command of human speech to distinguish, comprehend and respond appropriately. It depends on the voice as the research object, it permits the machine to naturally recognize and comprehend human spoken languages through speech signal processing and pattern recognition. The speech recognition technology is the cutting edge that permits the machine to transform the voice signal into the fitting content or command through the way toward distinguishing and comprehension. Speech recognition is cross-disciplinary and includes a wide reach. It has a very close relationship with acoustics, phonetics, semantics, information hypothesis, design recognition hypothesis and nervous system science disciplines. With the quick advancement of computer hardware and software and information technology, speech recognition technology is continuously turning into a vital technology in the computer information handling technology. The objective in automatic speech recognition is to give a way to verbal human-to-machine correspondence. Albeit both speech coding and recognition include analysis of the speech wave, the voice recognition issue is by a wide margin more troublesome. Applications of speech recognition technologies include process automation, telephone inquiry, automatic banking, secure voice access, to give some examples. In spite of the fact that the research in speech recognition is portion driven, there is a significant lack between research and commercial deployment. Understanding speech requires the coordination of various extraordinary and complex measures, for example, signal processing (recognition of phonemes, syllables and words), syntactic parsing and semantic analysis. Language is a framework that empowers a speaker to utilize words that are typically picked up during youth. The attributes of speech sounds rely upon the specific human language or tongue. Essentially speech recognition is a pattern recognition issue. Speech Recognition Systems are generally classified as discrete or continuous systems that are speaker dependent, independent or adaptive.

A speaker-dependent system requires that the user record an example of the word, sentence or phrase prior to its being recognized by the system i.e. the user trains the system. Some speaker-dependent systems require only that the user record a subset of system vocabulary to make the entire vocabulary recognizable. A speaker-independent system does not require any recording prior to system use. It is developed to operate for any speaker of a particular type. A speaker adaptive system is developed to adapt its operation to the characteristics of new speaker.

2. SPEECH RECOGNITION SYSTEM

The speech sound is captured using microphone to convert it in to electrical signal. The purpose of sound card inside the computer is to change analog signal into digital signal. Sound card has capabilities to store and play this speech signal.

Following are the building blocks for general speech recognition system.

A. Signal preprocessing B. Feature extraction

C. Language model D. Decoder E. Speech Recognition

2.1 Signal Pre-processing

Speech signal captured by microphone, telephone etc. are analog in nature so it is required to be digitized as per Nyquist theorem. This theorem states that a signal is to be sampled more than twice the rate of highest frequency present in it. Generally sampling frequencies for speech signal are 8 KHz and 20 KHz. For telephonic speech signal it is recommended to have 8 KHz sampling rate while 16 KHz is generally used for normal microphones.

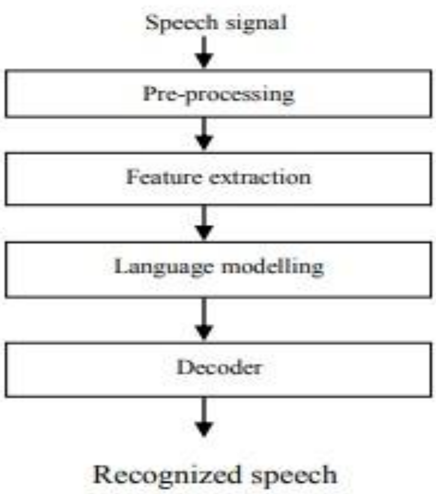

Fig -1: Basic working of Speech Recognition System

2.2 Feature Extraction

Feature extraction is used to find a set of properties that are stable and acoustically correlated to each other. So, it is a type of parameterization of speech signal. Such parameters can form the observation vectors. The goal of feature extractor is to identify relevant information for purpose of accurate classification.

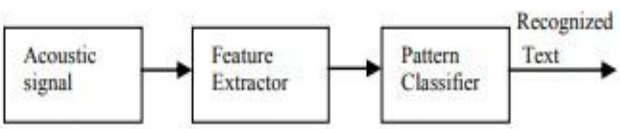

Fig -2: Role of Feature Extraction in Speech Recognition

Linear predictive Cepstral coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC) and PLP are some of the commonly used feature extraction techniques. The most widely used feature extraction techniques are discussed below:

2.2.1 Linear Predictive Coding (LPC)

It is a dominant, vigorous, perfect, consistent and trendy tool for voice recognition. It is linear combination of previous samples. The main purpose of LPC is frame-based examination of the input speech signal to produce experimental vectors. Each sample in LPC can be estimated as precedent samples in linear combination. To implement LPC and to produce the features, the input speech signals requires to surpass through pre emphasizer. The production of pre-emphasizer performs as the input to frame blocking where the signal is blocked into frames of N samples. In the next step windowing is done where each frame is windowed in such a manner to reduce signal disruption at the starting and end of each frame. After this step each windowed frame is auto correlated and the maximum autocorrelation value provides the order of LPC analysis and finally the resultant are LPC coefficients.

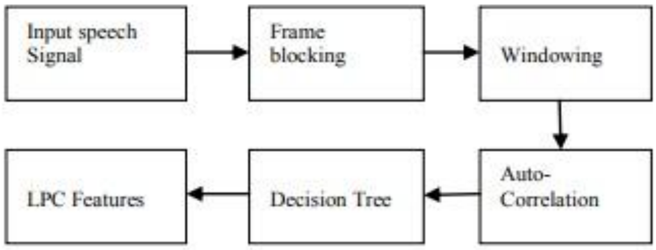

Fig -3: Steps to extract LPC Features

2.2.2 Perceptual Linear Prediction (PLP)

PLP model was given by Hermansky. It models the human speech which is based on the perception of psychophysics of hearing. It rejects inappropriate information of the speech and thus modifies the speech recognition rate. It is similar to LPC excluding its spectral features have been changed to match features of human auditory system. The PLP speech analysis method is more modified to human hearing, in comparison to the traditional Linear Prediction Coding (LPC).

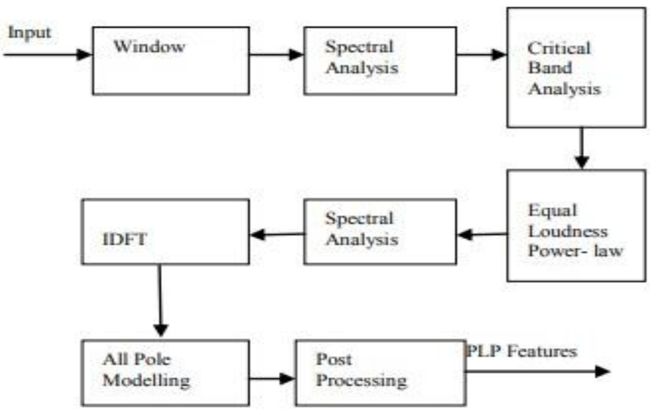

Fig -4: Block diagram of PLP feature extraction approach

2.2.3 Relative Spectral Filtering (RASTA)

This method was initially introduced with the purpose of dropping the unnecessary noise in Automatic speech recognition. The RASTA technique not only improves the effect of noise in speech signal but it also improves the worth of speech with surroundings noise. It also includes linear filtering of trajectory of power spectrum in the case of noisy speech. RASTA filter band passes each feature coefficient. This method can be improved by combining it with PLP for better performance.

2.2.4 Mel Frequency Cepstral Coefficient (MFCC)

Mel Frequency Cepstral Coefficient has a huge achievement in speaker recognition system. The MFCC is best acknowledged and most extensively used for both speech and speaker recognition. When the frequency bands are placed logarithmically in MFCC, it estimates the human system response more carefully than any other system. The method of processing MFCC is based on the short-term analysis, and thus from each frame a MFCC vector is computed. In order to obtain the coefficients, the speech samples are taken as the input and hamming window is applied to reduce the disruption of a signal. Then Discrete Fourier Transform (DFT) will be used to produce the Mel filter bank. MFCC can be calculated by using the formula -

Mel (f) = 2595*log10 (1+f/700).

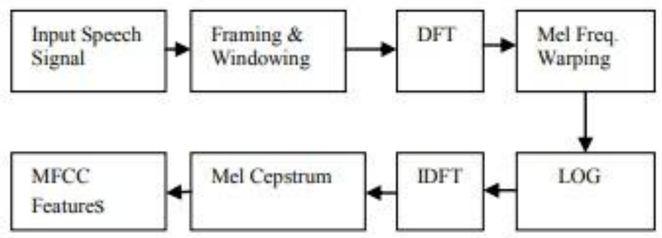

Fig -5: Steps involved in MFCC feature extraction

2.3 Language modeling

Language modeling is used to find the correct word sequence by predicting nth words using (n-1) preceding words. Language modeling is of various types.

- Uniform model: where occurrence of each word is equally probable.
- Stochastic model: probability of present word depends on probability of word preceding it.
- Finite state languages: This language modeling use finite state network to define allowed word sequence.

Context free grammar (CFG): It is used to encode allowed sequence of words in speech recognition system.
CFG follows a mechanism for defining languages (sets of acceptable sentences).
It is powerful as compared to Finite state grammars (FSG) due to imposing more structure on sentences. Human languages are actually close to the

- CFGs so it has application in speech recognizers. Generally, a CFG is defined by following relation,
- A finite set of terminal symbols (Defines words in vocabulary).
- A finite set of non-terminal symbols to define concepts.
- A special non-terminal symbol, represents the CFG.

Then a finite set of production rules are applied to detect a terminal or a non-terminal symbol in data sequence.
So, CFG is used to model a language in speech recognition by expanding its special non-terminal symbols after application of production rules.

2.4 Decoder

This stage is involved to find most likely word sequence for the given observation sequence. Generally dynamic programming algorithms are used to solve this problem. The purpose of these algorithms is to search single path through the network to have best match for the given sequence, Viterbi algorithm is mostly used for this purpose. In case of large vocabulary, a beam search method is useful for Viterbi iteration.

2.5 Speech Recognition

Speech recognition is completed in two phases: Training and testing phase. Training phase is just similar to identification of objects. It may be repeated many times for better recognition which improves the performance while testing. Testing phase includes the comparison between reference pattern scored while training, and spoken words at the time of testing. The extent of closeness in these two phases counts for improving the performance of the system. But variability encountered during recognition effects a lot for constant recognition rate.

3. CHARACTERISTICS OF SPEECH RECOGNITION SYSTEM

There are many variables in the systems of speech recognition and it is necessary to know these variables so that we can determine the algorithm appropriate to the system and the most important of these variables:

Types of Speech in most studies, speech is resumed into four types:

- **Isolated Words:** Isolated word recognizes attain usually require each utterance to have quiet on both sides of sample windows. It accepts single words or single utterances at a time. This is having “Listen and Non-Listen state”. Isolated utterance might be a better name of this class. Isolated word recognition-both speakers trained and speaker independent. This technology opened up a class of applications called ‘command and control’ applications in which the system was capable of recognizing a single word command (from a small vocabulary of single word commands), and appropriately responding to the recognized command. One key problem with this technology was the sensitivity to background noise (which were often recognized as spurious spoken words) and extraneous speech which was inadvertently spoken along with the command word. Various types of 'keyword spotting' algorithms evolved to solve these types of problems.

- **Connected Words:** The Connected word system is similar to isolated words, but allow the separate utterance to be "run together minimum pause between them. Connected word recognition-both speakers trained and speaker independent. This technology was built on top of word recognition technology, choosing to exploit the word models that were successful in isolated word recognition, and extend the model to recognize a concatenated sequence (a string) of such word models as a word string. This technology opened up a class of applications based on recognizing digit strings and alphanumeric strings, and led to a variety of systems for voice dialing, credit card authorization, directory assistance lookups, and catalog ordering.
- **Continuous Speech:** Continuous speech recognizers allows user to speak almost Naturally, while the computer determine the contents. Recognizer with continued speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries. Continuous or fluent speech recognition-both speakers trained and speaker independent. This technology led to the first large vocabulary recognition systems which were used to access databases (the DARPA Resource Management Task), to do constrained dialogue access to information (the DARPA ATIS Task), to handle very large vocabulary read the speech for dictation (the DARPA NAB Task), and eventually were used for desktop dictation systems for PC environments.
- **Spontaneous Speech:** At a basic level, it can be thought of as speech that is natural Sounding and not rehearsed. An ASR System with spontaneous speech ability should be able to handle a variety of natural speech feature such as words being run together. Spontaneous conversation systems which are able to both recognize the spoken material accurately and understand the meaning of the spoken material. Such systems, which are currently beyond the limits of the existing technology, will enable new services such as, 'Conversation Summarization', 'Business Meeting Notes', 'Topic Spotting' in fluent speech (e.g., from radio or TV broadcasts), and ultimately even language translation services between any pair of existing languages.

Size of the Vocabulary the volume of vocabulary used in the speech recognition system is important because it affects the complexity and processing requirements and determines the accuracy of the system. We note that there are applications that use only a few words, while others require the use of a very large number. There are no specific definitions, but we can define them 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.

Speaker Dependence:
• Speaker dependent system: Systems that require the user to train the system according to the user's voice.
• Speaker independent system: Systems developed for any speaker.
• Speaker adaptable system: Developed to adapt to the characteristics of new speakers.

4. MODELLING TECHNIQUES OF SPEECH RECOGNITION SYSTEM

The basic objective of modeling technique is to generate speaker models using speaker specific feature vector. The speaker modeling technique divided as: speaker recognition and speaker identification. It automatically identifies who is speaking on the basis of individual information integrated in speech signal. The speaker recognition is also divided into two parts that means speaker dependent and speaker independent. In Speaker independent approach of the speech recognition the computer should ignore the speaker specific characteristics of the speech signal and extract the expected message. Beside this, in speaker dependent, recognizing machine should extract speaker characteristics in the acoustic signal. The main aim of speaker identification is comparing a speech signal from an unknown speaker to a database of known speaker. The system can recognize the speaker, which has been trained with a number of speakers. Speaker recognition can also be divided into two methods, text-dependent and text independent methods. In text-dependent method, the speaker says key words or sentences having the same text for both training and recognition trials, whereas text independent does not dependent on a specific text being spoken. Following are approaches to speech recognition.

4.1 The Acoustic-Phonetic Approach:

The earliest approaches to speech recognition were based on finding speech sounds and providing appropriate labels to these sounds. This is the basis of the acoustic phonetic approach, which postulates that there exist finite, distinctive phonetic units (phonemes) in spoken language and that these units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time. Although, the acoustic properties of phonetic units are highly dangling, both with speakers and with neighboring sounds, it is assumed in the acoustic-phonetic approach that the rules governing the variability are straightforward and can be readily learned by a machine. The first step in the acoustic phonetic approach is a spectral analysis of the speech combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units. Segmentation and labeling phase done in next step. In this the speech signal is segmented into stable acoustic regions which followed by linking one or more phonetic labels to each segmented region. This results in a phoneme lattice characterization of the speech. In the last step it attempts to determine a valid word from the phonetic label sequences produced by the segmentation to labeling. Linguistic constraints on the task are preempted in order to access the lexicon for word decoding based on the phoneme lattice. This is called as validation process. The acoustic phonetic approach has not been widely used in most commercial applications.

4.2 Pattern Recognition Approach:

There are two essential steps involved in pattern recognition approach, pattern training and pattern comparison. Using a well formulated mathematical framework and initiates consistent speech pattern representation for reliable pattern comparison. A set of labeled training samples through formal training algorithm is essential feature of this approach [12]. In this, there exist two methods: Template base approach and stochastic approach. It is more suitable approach to speech recognition as it uses probabilistic models to deal with undetermined or incomplete information. There exist many methods in this approach like HMM, SVM, DTW, VQ etc., among these Hidden Markov Model is most popular stochastic approach today.

4.2.1 Template-Based Approach:

In template-based approaches matching, unknown speech is compared against a set of pre-recorded words in order to find the best match which is advantageous to find accurate word models. But it also has the disadvantage that pre-recorded templates are fixed. Due to this variation in speech can only be modelled by using many templates per word, which eventually becomes impractical. In this approach, the templates usually consist of representative sequences of features vectors for corresponding words. The fundamental aim here is to align the utterance to each of the template words and then select the word or word sequence that contains the perfect match. For each utterance, the distance between the template and the observed feature vectors are computed using various distance measure and these local distances are compile along each possible alignment path. Then the lowest scoring path identifies the optimal alignment for a word and the word template obtaining the lowest overall score interpret the recognized word or sequence of words.

4.2.2 Statistical-Based Approach:

In Statistical based approaches deviation in speech are modelled statistically. It uses automatic statistical learning procedure, typically the HMM. The approach represents the current state of the art. The main drawback of statistical models is that they must take priori modelling assumptions which are liable to be misleading, handicapping the system performance. In recent years, A new technique has been emerged which challenges the problem of conversational speech recognition has emerged. It is anticipated to overcome some fundamental limitations of the conventional Hidden Markov Model (HMM) approach. This new approach is a radical withdrawal from the current HMM-based statistical modeling approaches. Instead of using a large number of unstructured Gaussian mixture components to account for the huge variation in the observable acoustic data of highly coarticulated natural speech. The new speech model which has been developed to provide a rich structure for the partially observed dynamics in the domain of vocal-tract resonances.

4.3 The Artificial Intelligence Approach (Knowledge Based Approach):

It is combination of acoustic phonetic approach and pattern recognition approach. The artificial intelligence approach attempts to mechanize the recognition procedure according to the way a person applies its brilliance in visualizing, analyzing. And then finally making a decision on the measured acoustic features. The artificial intelligence approach attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analyzing, and finally making a decision on the measured acoustic features. Highly master system is used widely in this approach. In Knowledge based approaches: An expert knowledge about variations in speech is hand coded into a system. It is beneficial of explicit modelling variations in speech; but unfortunately, such expert knowledge is difficult to obtain and use successfully. Thus, this approach was judged to be impractical and automatic learning procedure was sought instead.

4.4 Connectionist Approaches (Artificial Neural Networks):

The artificial intelligence approach attempts to mechanize the artificial intelligence approach attempts to mechanize the recognition procedure according to the way a person applies its intelligence in visualizing, analyzing, and finally making a decision on the measured acoustic features. Among the techniques used within this class of methods are uses of an expert system that integrates lexical, phonemic, semantic, syntactic and pragmatic knowledge for segmentation and labeling. It uses tools such as artificial NEURAL NETWORKS for learning the relationships among phonetic events. It mainly focuses on the representation of knowledge and integration of knowledge sources. This method has not been widely used in commercial systems.

4.5 Vector Machine (SVM):

One of the powerful tools for pattern recognition is SVM which uses a discriminative approach is a SVM. It uses linear and nonlinear separating hyper-planes for data classification. In spite of this, SVMs can only classify fixed length data vectors. This procedure may not be readily applied to task involving variable length data classification. This data has to be trans-formed to fixed length vectors before SVMs can be used. It is a generalized linear classifier with maximum-margin fitting functions. This function gives regularization which helps the classifier generalized better. Traditional statistical and Neural Network methods control model complexity by using a small number of features. SVM curbs the model complexity by con-trolling the VC dimensions of its model. This method is independent of dimensionality and can employ spaces of very large dimensions spaces, which grants a construction of very large number of non-linear features and then performing robust feature selection during training.

5. PERFORMANCE EVALUATION OF SPEECH RECOGNITION SYSTEMS

The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy which is usually rated with Word Error Rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR) [24]. Word Error Rate (WER): Word error rate is a common metric of the performance of a speech recognition or machine translation system. The general difficulty of measuring performance lies in the fact that the recognized word sequence can have a different length from the reference word sequence. The WER is derived from the Levenshtein distance, working at the word level instead of the phoneme level. This problem is solved by first aligning the recognized word sequence with the reference (spoken) word sequence using dynamic string alignment. Word error rate can then be computed as

$$WER = (S+D+I)/N$$

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

When reporting the performance of a speech recognition system, sometimes Word Recognition Rate (WRR) is used instead:

$$WRR = 1 - WER = 1 - (S+D+I) / N = (H - I) / N$$

Where H = (N-S-D) is the correctly recognized words

5. CONCLUSIONS

Voice recognition is computer analysis of the human voice, particularly for the target of translating words and phrases and routinely identifying who is speaking on the foundation of individual information incorporated in speech waves. This procedure makes it feasible by using the voice of presenter and it is easy to authenticate their individuality. In this paper, we have presented a review of Automatic Speech Recognition Systems. In the second section, we discussed in detail the basic building blocks of a Speech Recognition System. We have also discussed the commonly used feature extraction techniques which contributes maximum recognition accuracy in any speech recognition application. In the third section, we have talked through the essential characteristics that are considered while determining the techniques that are to be used. Depending on the application of the system, the most important characteristic is determined and the System is built around that. In the fourth section we have studied different approaches and modelling techniques thar are most widely used and their applications. These include techniques like Acoustic-Phonetic approach, Pattern Recognition approach, AI Networks, SVMs, etc. And, in the last section we have understood how the performance of a Speech Recognition System is evaluated.

REFERENCES

[1] S. K. Gaikwad et.al, “A review on speech recognition technique”, International journal of Computer Applications, vol. 10, no. 3, November 2010.

[2] M. A. Anusuya and S. K. Katti, “Speech Recognition by Machine: A Review”, International journal of Computer Science and Information Security, vol. 6, no. 3, 2009, pp.181-205.

[3] S. Preeti, K. Parneet, “Automatic Speech Recognition: A Review”, International journal of engineering trends and technology, vol. 41, Issue 2, Haryana India, 2013.

[4] M. Agarwal and T. Raikhwar, “Speech recognition using signal processing techniques”, International journal of engineering and innovative technology (IJEIT), vol. 5, Issue 8, February 2016.

[5] K. A. Qazi, T. Nawaz et.al, “A hybrid technique for speech segregation and classification using a sophisticated deep neural network”, PLoS ONE 13(3): e0194151, March 20, 2018, pp. 1-15.

[6] Dominique, F., Odile, M., Irina, I. (2017). New Paradigm in Speech Recognition: Deep Neural Networks, the ContNomina project supported, French National Research Agency (ANR).

[7] Vrinda1, Shekhar, Chander. (2013). Speech Recognition System for English Language, International Journalof Advanced Research in Computer and Communication Engineering, 2(1), January 2013, ISSN (Print): 2319-5940 ISSN (Online): 2278-1021, www.ijarcce.com, Copyright to IJARCCE.

[8] Santosh, K. G., Bharti, W. G., Pravin, Y. (2010). A Review on Speech Recognition Technique, International Journal of Computer Applications (0975 – 8887), 10(3).

[9] Ashok Kumar, Vikas Mittal. Y. (2019). Speech Recognition: A Complete Perspective. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019

[10] E. Chandra and C. “A review on Speech and Speaker Authentication System using Voice Signal feature selection and Extraction”, 2009 IEEE International

Advance Computing Conference (IACC2009) Patiala, India, 6-7 March 2009.

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

[12] Karpagavalli, S., & Chandra, E. (2016). A Review on Automatic Speech Recognition Architecture and Approaches. International Journal of Signal Processing, Image Processing and Pattern Recognition, 9(4), 393-404.

[13] Sapijaszko, G. I., & Mikhael, W. B. (2012, August). An overview of recent window-based feature extraction algorithms for speaker recognition. In Circuits and Systems (MWSCAS), 2012 IEEE 55th International Midwest Symposium on (pp. 880-883). IEEE.

[14] Jung, Y. G., Han, M. S., & Lee, S. J. (2007). Development of an optimized feature extraction algorithm for throat signal analysis. ETRI journal, 29(3), 292-299.

[15] Dave, N. (2013). Feature extraction methods LPC, PLP and MFCC in speech recognition. International journal for advance research in engineering and technology, 1(6), 1-4.

[16] Kumar, J., Prabhakar, O. P., & Sahu, N. K. (2014). Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review. International Journal of Innovative Research in Computer and Communication Engineering, 2(1), 2760-2269.

[17] Singh, A., Panchal, T., & Saharan, M. (2013). Review on Automatic Speaker Recognition System. International Journal of Advanced Research in Computer Science and Software Engineering, 3(2), 350-354.

[18] Sunitha, C., & Chandra, E. (2015). Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm. International Journal of Engineering and Technology (IJET), 7(5), 1685-1692.

[19] Bhupinder Singh, Neha Kapur, Puneet Kaur "Sppech Recognition with Hidden Markov Model:A Review" International Journal of Advanced Research in Computer and Software Engineering, Vol. 2, Issue 3, March 2012.

[20] Shivanker Dev Dhingra, Geeta Nijhawan, Poonam Pandit, "Isolated Speech Recognition using MFCC and DTW" International Journal of Advance Research in Electrical, Electronics and Instrumentation Engineering, Vol.2, Issue 8, August 2013.

[21] Ibrahim Patel, Dr. Y. Srinivas Rao, "Speech Recognition using HMM with MFCC-an analysis using Frequency Spectral Decomposition Technique" Signal and Image Processing:An International Journal(SIPIJ), Vol.1, Number.2, December 2010.