A Study of Speech Recognition

Abhi J. Patel
Dept Master of computer Applications,
Parul University
Vadodra, India

Akhil Patel
Dept Master of computerApplications,
Parul university,
Vadodra, India.

Prof Kaushal Gor
Dept Master of computerApplications,
Parul university,
Vadodra, India.

Abstract

To be able to control devices by voice had always intrigued mankind. Today after intense research, Speech Recognition System, has made a niched for themselves and can be sown in many walks of life. The accuracy of Speech Recognition System remained one of the most important research challenges e.g. noises, speaker variability, language variability, vocabulary size and domain. The design of speech recognition system required careful attentions to the challenging such as various types of Speech Classes and Speech Representations, Speech Pre-processing stages, Feature Extraction techniques, Database and Performance evaluations. This paper present the advances made as well as highlights the pressing problems for a speech recognition system. The paper also classified the system into Front-End and Back-End for better understanding and representation of speech recognition system in each part.

Keywords

Acoustic model, Feature Extraction, Hidden Markov Model, Neural Networks.

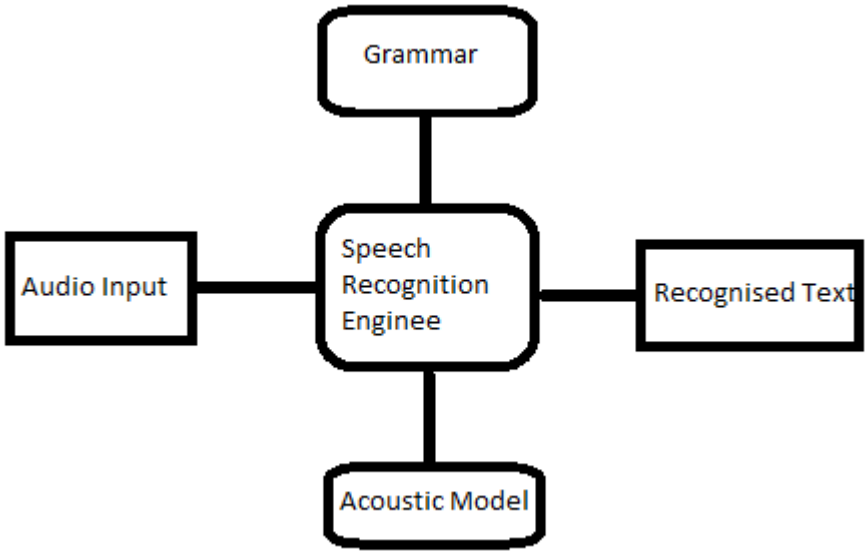
Introduction

Since ages speech has been a simple mean of communication between humans. Speech Recognition is the process of changing associate degree acoustic speech into text, and / or identification of the speaker.

A system engineered at Bell Laboratory in 1952 that was the primary word recognition system that was trained to acknowledge digits [3]. Some of the wide used speech recognition systems are sorts of Speech Recognition Systems. Some of Speaker Dependent Systems, Speaker freelance System, Isolated Word Recognizer, Connected Word Recognizer, and Spontaneous Recognition System.

Over the years the Speech Recognition Systems have come back an extended approach the method has ensured its presence because of the well-established would like of voice operated systems. However, there is a lot to be accomplished. Most of analysis done to this point could be attributed to terribly fact the actual fact} that speech is a very subjective development. The general identified issues are Speaker Variation, background signal and Continuous Character of Speech. Perhaps the foremost evident supply of performance degradation in speech recognition is Noise. Noise can beclassified as either environmental i.e. traffic, rain, people talking or speaker enclosed i.e. coughing, sneezing, swallowing, breathing, chewing, etc.

In this article, Speech Recognition System has been subdivided into Front-End and Back-End (as shown in Figure 1 below), based on the subdivision a brief review of work done so far within the domain of speech recognition system has been given.

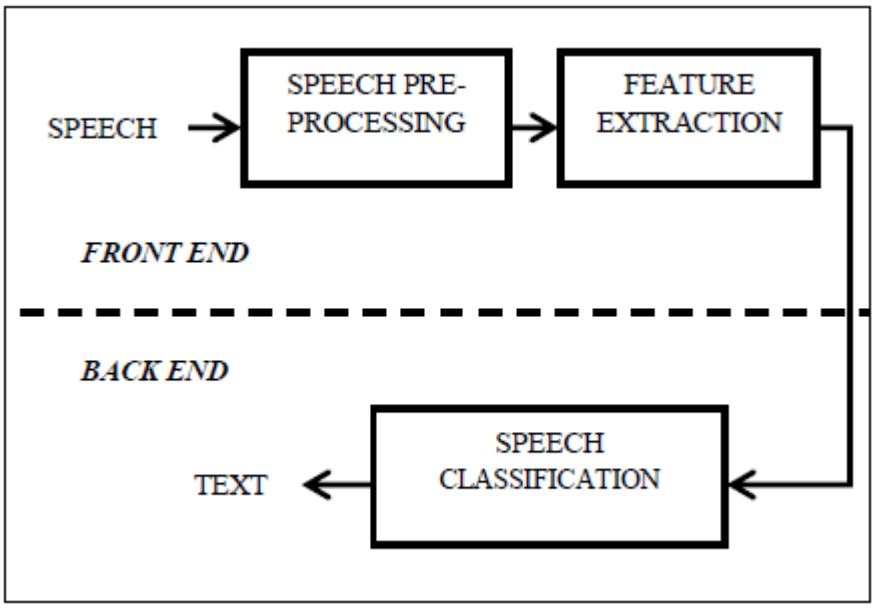


Grammar

Language model just defines the possible sequences of words and their probabilities. Grammar is simply another sort of the language model that isn't trained however written manually. Grammar usually defines simple subset of the language. For a lot of complicated set of the language applied math models area unit easier to make, though grammars could work too. In most systems each grammars and applied math language models area unit regenerate to a standardized format (WFST)

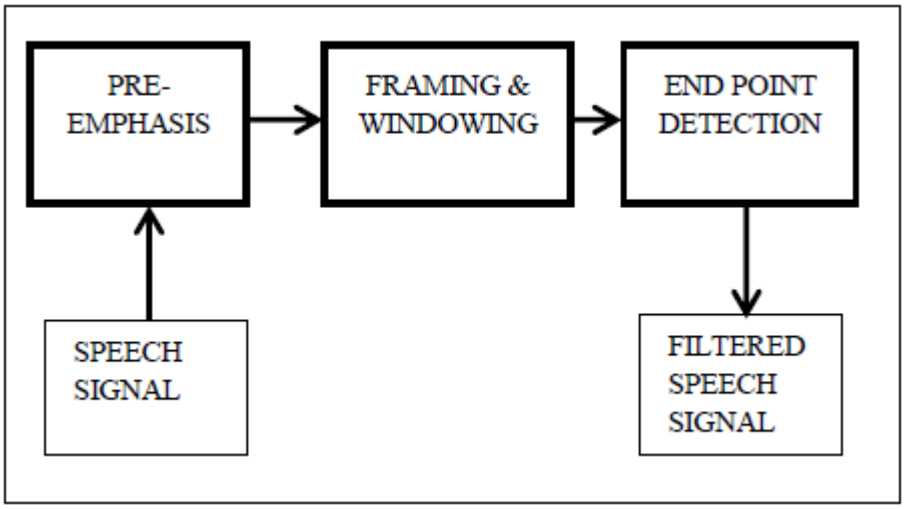
Acoustic model

An acoustic model's task is to compute the P(O|W), i.e. the chance of generating a speech wave for the mode. An acoustic model, as a very important a part of the ASR system, accounts for a large part of the computational overhead and also determines the system's performance. GMM-HMM-based acoustic models are widely used in traditional speech recognition systems. In this model, GMM is used to model the distribution of the acoustic characteristics of speech and HMM is used to model the time sequence of speech signals. Since the increase of deep learning in 2006, deep neural networks (DNNs) are applied in speech acoustic models. In 2009, Hinton and his students used feed forward fully-connected deep neural networks in speech recognition acoustic modelling[1].



Front-End Analysis

Front-End of the speech recognition system includes of Speech Preprocessing and have Extraction Block. Noise and variations in Amplitude of the signal will hardly influence the integrity of a word whereas temporal arrangement variations will cause an oversized distinction amongst samples of a similar word. These problems are proscribed within the Signal Preprocessing half. Preprocessing generally involves End Point Detection, Pre-emphasis Filtering, Noise Filtering, Framing, Windowing, Echo Cancelling, etc. [4]. Block Diagram for Signal Preprocessing stage is shown in Figure two below.



Feature Extraction may be a method extracting specific options of the preprocessed speech signal. This can be through with various forms of Techniques like Cepstrum analysis, pic, MFCC (Mel Frequency Cepstrum Coefficient), LPC (Linear prognosticative Coefficient), etc. In 1976, L.R

Back-End Analysis

Back-End consists of Speech Classification block. Speech Classification process is for classifying the extracted features and relates the input sound to the best fitting sound from a database and represents them as an output. The commonly used techniques for Speech Classification are HMM (Hidden Markov Model), DTW (Dynamic Time Warping), VQ (Vector Quantization), ANN (Artificial Neural Network), etc. [3]. In many Speech recognition systems, hybrid techniques are implemented and work in a cooperative relationship. Neural Networks perform very well at learning phoneme probability from highly parallel audio input, while Markov Models can use the phoneme observation probabilities that Neural Networks provide to produce the likeliest phoneme sequence or word.

In [20], Comparison of two different types of Neural Networks i.e. Multi-Layer Feed Forward and Radial Basis Function Network for Speech Recognition when Mel- Frequency Cepstrum Coefficient is used in Signal Preprocessing stage. Here RBF network needs more amount of Hidden Layer as compared to Multi-Layer Feed Forward Network and increase in Number of Hidden Layer increases Computational time of system. In 2014, Amr Rashed gives comparative analysis of different Neural Network Learning Algorithms [21]. A Feed Forward Multi-Layer Perceptron Neural Network algorithm gives fast and accurate result even in presence of Whit Gaussian Noise. Here we can also observe that Sequential Weight/Bias training algorithm gives efficient result in Speech Recognition Systems.

T. Lee, P.C. Ching and Lai-Wan Chan propose a novel approach of utilizing Recurrent Neural Network (RNN) for Isolated Word Recognition [22]. Here the RNN Speech Model is trained in two stages. First, The RSM‟s are trained independently to extract the Temporal and Static characteristic of individual words. Second, Mutual discriminative training among the RSM‟s takes place for minimizing the probability of misclassification and improving the recognition accuracy. In 2012, K. Dutta and K.K. Sharma proposed a combined architecture of LPC and MFCC Feature Extraction technique by two different RNN. This combined Architecture gives 10% gain in recognition rate than individual architectures. Hence the result obtained by the proposed system can be easily improved by using Hidden Layered based RNN [23]. In

Parallel they have also proposed a Dynamic Segmentation of Voiced/Unvoiced segments from speech utterance [24]. This results in 90% Recognition Rate but the Testing Time of the system is increased by small amount which can be counterbalance byusing parallel processing technique results in improvement of speed of computation.

for Speech Recognition of Hindi Hybrid Paired words and observes that Consonant dominated words provides better Recognition rate as compared to Vowel dominated words [25]. In [4], Comparative analysis of different training algorithm is presented in which "trainscg" training algorithm performs well over a wide variety of problems. Here for efficient Speech Recognition System Neural Network with MFCC is used. The result can be improved by increasing the training data size. In 2013, Manan Vyas designed a GMM based Speaker-Dependent Speech Recognition System [26]. In this System End Point of speech utterance is detected by concept of Energy & ZCR and MFCC is employed as feature extraction technique. Hence GMM gives a poor Recognition Rate (70%) as compared to alternative classifiers, however just in case of Speaker Recognition it gives efficient results.

Conclusion

Speech Recognition may be a difficult drawback to manage. We have tried during this paper to produce a review of however much this technology has progressed within the previous years. The performance of Speech Recognition System is mainly depends on the quality of Signal Preprocessing Stage. The Preprocessing quality is giving the largest impact on the Speech Classification performance. Signal Preprocessing consist an EPD, Filtering, Framing, Windowing, Echo Cancellation, etc. An Improvement in any individual part can improve the overall system performance. For effective working of Back-End there ought to be additional efforts in Front- End processing. MFCC is more preferred in Feature Extraction technique because it generates the coaching vectors by transforming speech signal into frequency domain, and therefore it is less affected by noise.

References

1. https://www.edx.org/course/speech-recognition-systems-3
2. https://searchcrm.techtarget.com/definition/speech-recognition
3. https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speech-recognition.htm