Advanced Research in Electrical and Electronic Engineering
p-ISSN: 2349-5804; e-ISSN: 2349-5812 Volume 5, Issue 2 April-June, 2018, pp. 119-122
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/Publication.html

Comparative Analysis of Speech Recognition Modelling Techniques

Shipra J. Arora[1] and Dr. Rishipal Singh[2]

[1,2]CSE Department, GJUST, Hisar
E-mail: [1]jkshipra22@gmail.com, [2]pal_rishi@yahoo.com

Abstract—*Speech is one of the most widely used communication tool among human beings. Now days, Speech recognition has been implemented in real world language applications. But most important challenge for researchers is accuracy of speech. There are various modelling techniques for speech recognition. The paper signifies the comparative analysis of various speech recognition techniques along with their relative merits and demerits.*

1. INTRODUCTION

Digital signal processing has emerged as a recognized discipline over the past three decades. Much of the thrust for this is representation, coding and reproduction of speech information. Speech is the most important aspect of communication among human beings. A lot of work has been done in the past in the area of speech processing but it has a wide variety of applications such as speech recognition, speaker identification, speech synthesis, machine translation, information retrieval system and others. Now these days, speech processing work is going on for all 22 Indian languages. The goal of researchers is to design a system that will investigate and understand human natural speech with great accuracy.

2. SPEECH RECOGNITION TECHNIQUES

There are three main approaches to speech recognition i.e. Acoustic Phonetic Approach, Artificial Intelligence Approach and Pattern Recognition Approach.

2.1 Acoustic Phonetic Approach

One of the earliest approaches to speech recognition is Acoustic Phonetic Approach. The objective of this approach was to find out sounds of speech and proper labels were provided to these sounds. Hemdal and Hughes in 1967 developed Acoustic phonetic approach. They described that there exists finite, distinctive phonetic units which are known as phonemes in the spoken language and a set of acoustics properties characterize these units that are manifested in the speech signal over time. Even though, these acoustic properties of phonetic units are vastly erratic both with speakers and neighbouring sounds which is so-called co-articulation effect. In Acoustic Phonetic approach, the first step is a spectral analysis of the speech. The most general spectral analysis techniques are the filter bank methods and the linear predictive coding methods. The next step is a feature detection which converts the spectral dimensions to the features set that illustrate the broad acoustic properties of the different Phonetic units. The next step is a segmentation and labelling phase in which the signal of speech is sub-divided into discrete regions and according to acoustic properties of signal, assigning phonetic labels to each segmented region. The last step in this approach is to determine a valid word or string of words from the phonetic label chains generate by the segmentation to labelling.

2.2 Pattern Recognition Approach

Rabiner and Juang in 1993 developed the pattern-matching approach which can be seen as a classification process. Its ultimate goal is to optimally extract patterns based on definite conditions and to differentiate one class of patterns from the other classes. It consists of four steps i.e. feature measurement, pattern training, pattern classification and decision logic. The essential feature of this approach is that it uses a well formulated mathematical framework. A succession of measurements is done through the input signal to define the test pattern. In order to create a reference pattern, one or more test patterns referring to speech sounds of the same class are taken. A speech template or a statistical model (e.g., a HIDDEN MARKOV MODEL or HMM) are the possibilities which can be applied to a sound smaller than a word, a word or a phrase. The unknown test pattern is compared with class reference pattern and calculation of distance measurement between the two is done in the pattern classification stage. The identity of the unknown test pattern according to the match reliability of the patterns is determined in Decision logic stage. In the past six decades, The pattern-matching approach has turned out to be the predominant method for speech recognition. This approach is shown in Figure 1 through a block diagram. Two methods existing are namely template method and stochastic model method.

2.3 Template Matching Method

In order to recognize the best match, in this technique, unknown pattern of speech is evaluated with a set of template patterns. It has benefit of using absolutely accurate word model. Speech recognition Template based method have granted various similar techniques that have advanced the field significantly in the past six decades. The underlying scheme is simple and shown in Figure 2. Reference pattern consists of a group of exemplary speech patterns which represents the glossary of candidate's words. For recognition, an unidentified spoken utterance is matched with each of these reference templates and choosing the best matching pattern category. Usually templates for entire words are constructed.

2.4 Stochastic Method

The usage of probabilistic models for dealing incomplete information is described in Stochastic methods. Deficiency arises from a variety of sources in speech recognition, for example speaker changeability, contextual effects and others. Thus, stochastic models provide appropriate approach to speech recognition. These days, hidden Markov modelling (HMM) is considered as the most admired stochastic approach.

A hidden Markov model is represented by a finite state markov model and a set of output distributions. In the Markov chain models, the transition parameters represent temporal variability's, whereas in the output distribution model,the parameters represent spectral variability's. These two categories of variability's are the essence of speech recognition process.

2.5 Dynamic Time Warping

DTW is a kind of template matching technique that finds the optimal alignment between two time series if one time series may be warped non-linearly by stretching or shrinking it along its time axis. The warping between two time series can then be used to find the similarity between the two time series. In speech Recognition, it is used to determine if two waveforms represent the same spoken phrase. In a speech waveform, each spoken sound duration and the interval between sounds are tolerable to fluctuate but the on the whole speech waveform must be similar. For example, one can detect walking patterns similarities easily if a person was walking slowly in one video and same person was walking quickly in another video. Dynamic Time Warping has been applied to audio, video and graphics. It can analyse any data which can be easily converted into a linear representation. Speech recognition which is coping with different speeds of speech is one of the well known application. As compared to other pattern matching algorithms, continuity is less important in DTW.

2.6 Vector Quantization

Vector Quantization is a technique which encodes an input vector into an integer that is linked with reproduction vectors codebook . The vector which is close to input vector has been chosen as the reproduction vector. The efficiency of coding ranges from 1 to N where N is the size of the codebook. For IWR, each word of the vocabulary acquire its own Vector Quantization codebook which is based on training progression of numerous repetitions of the word. All codebooks assess the test speech and whose codebook yields the lowest distance measure, ASR chooses that word.

2.7 Artificial Intelligence Approach

Fusion of both the acoustic phonetic approach and pattern recognition approach is the AI approach. AI approach automates the speech recognition procedure in visualizing, analysing and making a decision on the features of measured acoustics according the manner in which a person applies his/her intelligence on it. Knowledge based method uses the information regarding spectrogram, phonetic and linguistic.

Feature Measurement

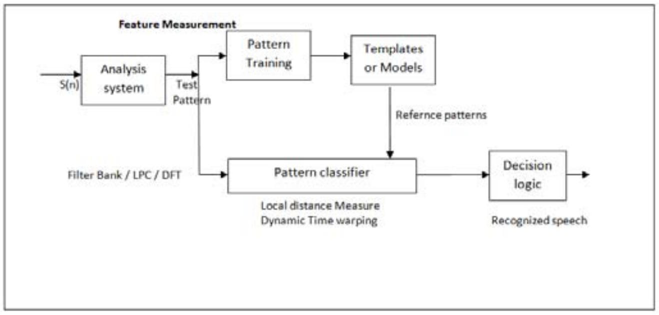

Fig. 1 Pattern Recognition Approach

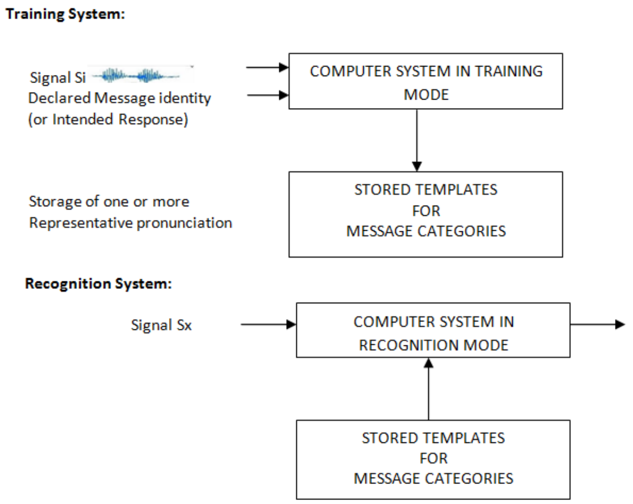
Training System:
Signal Si
Declared Message identity
(or Intended Response)
COMPUTER SYSTEM IN TRAINING MODE
Storage of one or more
Representative pronunciation
STORED TEMPLATES FOR MESSAGE CATEGORIES
Recognition System:
Signal Sx
COMPUTER SYSTEM IN RECOGNITION MODE
STORED TEMPLATES FOR MESSAGE CATEGORIES

Fig. 2 Template Matching Method

3. COMPARATIVE ANALYSIS

Acoustic Phonetic Approach has several difficulties as follows

a) One must have widespread awareness of acoustic properties of phonetic units.

b) To tune labelled speech, there is no well defined procedure.

c) To label training speech, there is no standard method.
d) Features are often based on non-optimal considerations

In order to successfully implement in real speech recognition system, this approach still requires much more awareness due to these limitations. It has not been widely used in most commercial applications.

The major advantage of Template matching method is that due to categorization of phonemes, it has avoided errors. Each isolated word must have its own reference template in turn. As vocabulary size increases, it becomes very expensive to prepare a template and its matching process. In template method, one key solution is to obtain typical speech frames sequences for a pattern (a word) through some procedure and to depend on the utilization of local spectral distance measures to compare patterns. Another key solution is to apply some variety of dynamic programming to align patterns momentarily to differentiate in speaking rates across speakers and across word repetitions by the same speaker. This provides good recognition performance for a variety of practical applications. But disadvantages of this method is that many templates per word are used to model the speech as pre-recorded templates are fixed, which eventually becomes unfeasible.

The strengths and weaknesses of the pattern recognition are as follows:

a) The system performance depends on the amount of training data used to create reference patterns. As training data is more, the higher the performance of the system.
b) Speaking environment and medium transmission characteristics which are used to create speech also effects reference patterns.
c) For both pattern training and pattern classification, computational load is directly proportional to pattern quantity which is being trained or recognized. So, for large number of sound classes, computation is prohibited.
d) This technique is not sensitive to the syntax and semantics of task and choice of vocabulary words. So basic techniques are applied to words and sub- words with slight algorithm modification.
e) It is relatively simple to include syntactic and semantic constraints directly into this approach and resulting in improving recognition accuracy.
f) As compared to the Stochastic approach, the template based approach is limited to word unit. It is very difficult to segment a sound smaller than a word but statistical approach is applied to both word and sub-word unit and has its advantages in large vocabulary system.

Hidden Markov Modelling (HMM) is more general as compared to template matching approach. It has a firm mathematical foundation. The fundamental problems for HMM design are a) the evaluation of the probability of a sequence of observations given a specific HMM. b) The determination of a best sequence of modal states and c) the adjustment of modal parameters so as to best account for the observed signal. Once these fundamental problems are solved, we can apply HMMs to selected problems in speech recognition.

DTW is an algorithm particularly suited to matching sequences with missing information, provided there are long enough segments for matching to occur. The optimization process is performed using dynamic programming, hence the name.

Advantages of vector quantization technique are reduced storage, reduced computation and efficient representation of speech sound, language independent, easy to train. Disadvantages are Limited number of templates, speaker specific and need actual training examples.

Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based methods provided little insight about human speech processing, thereby making error analysis and knowledge-based system enhancement difficult but these methods have been very effective in the design of a variety of speech recognition systems. On the other hand, linguistic and phonetic literature provided insights and understanding to human speech processing. However, this approach had only limited success, largely due to the difficulty in quantifying expert knowledge. Another difficulty is the integration of levels of human knowledge phonetics, lexical access, syntax, semantics and pragmatics. knowledge also has also been employed to lead the models design and algorithms of various different techniques such as template method and stochastic modelling. This type of knowledge makes a significant distinction between knowledge and algorithms. Algorithms enable us to solve problems. Knowledge enables the algorithms to work better. This form of knowledge based system enhancement has contributed considerably to the design of all successful strategies reported.

4. CONCLUSIONS

Various techniques for speech recognition has been studied in the present paper. It provides comparative analysis among various speech recognition modelling techniques. By comparing and contrasting, it has been found that Hidden Markov Modelling techniques is well suited to speech recognition.

REFERNECES

[1] Kadyan, Virender, Archana Mantri and R.K.Aggarwal, “Refinement of HMM Model Parameters for Punjabi Automatic Speech Recognition (PASR) System” IETE Journal of Research (2017), pp 1-16

[2] Harpreet Kaur & Rekha Bhatia, “Speech Recognition System for Punjabi Language”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, Issue 8, ISSN: 2277-128X, Aug. 2015.

[3] Arora Shipra J. and, Singh Rishipal; “Acoustic and Phonological Analysis of Homophones of Punjabi Language” International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) ISSN (P): 2249-6831; ISSN (E): 2249-7943 Vol. 4, Issue 1, Feb 2014, pp 95-102

[4] Arora Shipra J. and, Singh Rishipal; “Automatic Speech Recognition: A Review” International Journal of Computer Applications (0975 – 8887) Volume 60– No.9, Dec 2012, pp 34-44

[5] Dhanjal Surinder and Bhatia S.S.; “A New Corpus for the Punjabi Speech processing”; International Symposium on frontiers of research on Music and Speech(FRSM-2012), KIIT Gurgaon, India; January 18-19, 2012, pp 223-227.

[6] Bhaskararao Peri; “Sailent phonetic features of Indian Languages in Speech Technology”; Sadhana Academy Proceedings in Engineering Sciences; Indian Academy of Sciences, Banglore, India; Volume 36, Number 5, October 2011, pp 587-599.

[7]. Sadaoki Furui, “50 years of Progress in speech and Speaker Recognition Research”, ECTI Transactions on Computer and Information Technology, Vol.1, No..2 ,November 2005.

[8].B.H.Juang and S.Furui, “Automatic speech recognition and understanding: A first step toward natural human machine communication”, Proc. IEEE, 88, 8, pp.1142-1165, 2000.

[9] C.H.Lee et.al., “Acoustic modelling for large vocabulary speech recognition”, Computer Speech and Language, vol. 4, pp. 127-165, 1990.

[10].B.Lowrre, “The HARPY speech understanding system, Trends in Speech Recognition”, W.Lea, Ed., Speech Science Pub., pp.576-586, 1990.

[11].L.R.Rabiner, ”A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proc. IEEE, 77(2):257-286, February 1989.

[12].K.P.Li and G.W.Hughes, “Talker differences as they appear in correlation matrices of continuous speech spectra”, J.Acoust.Soc.Am., 55, pp.833-837, 1974.