A Study on Automatic Speech Recognition

Sunilkumar M Hattaraki
Department of Electronics and Communication Engineering
B.L.D.E.A's V. P. Dr. P. G. Halakatti College of
Engineering and Technology,
Vijayapura, Karnataka, India.

Vinodakumar M.Sonar
Department of Electronics and Communication Engineering
Government Polytechnic College ,
Bagalkote, Karnataka, India.

Abstract:- Over the last few decades, speech signal processing has been the most interesting and broad area of research. Speech is an effective means of communication for humans to converse with one another. Speech can be heard by human ears in a real-world environment, even with all of the background noises. Distinguishing the target speech from the background noise by a machine, on the other hand, remains a difficult issue. Man has always had a strong desire to make machines more human-like. This desire has resulted in the development of disciplines such as artificial intelligence (AI), which mimic human behavior in machines.

Keywords:- Speech Signal, Machine, Recognition, Noise and Artificial Intelligence (AI).

I. INTRODUCTION

A. Feature Extraction Technique

Pattern recognition is a subset of speech recognition. In supervised pattern recognition, there are two stages: training and testing. In both phases, the process of extracting features relevant for classification is the same. The classification model's parameters are estimated using a large number of class instances during the training phase (Training Data). Each class's trained model is matched with a feature of the test pattern during the testing or recognition phase[1] (Test Speech Data). The model that best matches the test pattern is proclaimed to be the test pattern's home.

B. Types of Speech Recognition

Based on the types of utterances they can distinguish, speech recognition systems can be divided into a variety of categories.

➢ *Isolated Word*

On both sides of the sample window, each voice should normally leave a separate word recognition feature. This does not imply that a single word is not supported, but rather that only one utterance is required at a time. A "listen and non-listen state" exists in this. "Isolated Utterance" is another term for this class. This is fine when the user just needs to offer one-word responses or orders, but it is strange when more words are required. The key advantages of this style are that word boundaries are visible and the words are usually specified clearly, making it very easy to perform. The results are biased by the boundaries chosen[1], which is a downside of this strategy.

➢ *Word that is Linked*

Like isolated words, connected word systems allow different utterances to be "run-to gether" with only a brief gap between them [2].

➢ *Continuous Speech*

Users can speak almost normally as the computer uses continuous speech recognizers to figure out what they're saying. It's essentially computer dictation. Continuous speech recognizers are among the most challenging to develop since they rely on complex algorithms to establish utterance boundaries. As one's vocabulary expands, it becomes more difficult to distinguish between distinct word sequences.

➢ *Impromptu Speech*

This is an unplanned and impromptu form of speech. Run-on words, "ums" and "ahs," and even moderate stutters should be handled by an ASR system that uses spontaneous speech. Mispronunciations, false starts, and non-words are all examples of unrehearsed or spontaneous speaking.

II. TECHNIQUES FOR RECOGNIZING SPEECH

The speaker recognition system consists of three steps. After accepting speech input, the system executes pre-processing activities before analysing the data. In order to construct the consequent string output depicted in Fig 1, feature extraction and classification (modelling) of the speech input are also carried out.

Fig.1. Speech Recognition Process

A. Analysis of Speech

Noise causes the most interference during voice recognition, which further decreases the system's performance. To produce a better result , it is necessary to remove the noise from the voice signal. before delivering it to the feature extraction block. This is accomplished through the use of pre-processing.

B. Speech Feature Extraction

The extraction of speech features is an important step that involves converting speech signals into a stream of feature vector coefficients that only contain the data required to identify a specific utterance. Because each speech contains

a unique set of features, a variety of feature extraction algorithms can be used to obtain and use these traits for speech recognition. However, when dealing with voice signals, extracted speech features must meet certain criteria, such as being easy to calculate, stable over time, and resistant to noise and surrounds.

C. Modeling of speech

➢ *Acoustic-Phonetic Methodology*

It is founded on the principle of acoustic phonetics, which argues that spoken language has finite, distinct phonetic units. The phonetic units are a set of acoustic qualities that reveal themselves over time in the speech signal, or spectrum. As illustrated in Fig.2, the acoustic phonetic technique starts with a speech spectrum analysis and then moves on to feature detection, which converts the spectral observations into a set of features that specify the general acoustic qualities of the individual phonetic units.
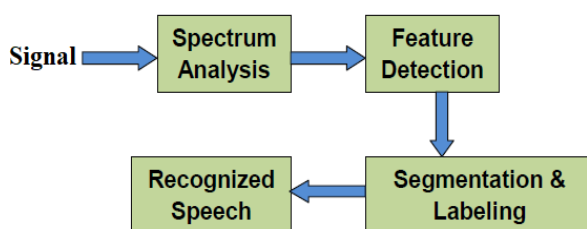

Fig.2. Speech Recognition Block Diagram (Acoustic-Phonetic Approach)

➢ *An Approach to Recognizing Patterns*

The pattern recognition method requires both pattern training and pattern comparison. This method is based on a mathematical framework that applies a formal training procedure to generate a consistent speech pattern representation from a set of labelled training samples for pattern comparison that is dependable. Both the template-based and stochastic approaches [3] are valid possibilities. Because stochastic models use probabilistic models to deal with indeterminate or partial information, they are excellent for voice recognition. HMM, SVM, DTW, VQ [4], and other stochastic approaches exist; nevertheless, the hidden markov model is the most widely used stochastic strategy today.

➢ *Artificial Intelligence-based Strategy*

Artificial Intelligence (AI) technique combines the auditory phonetic and pattern recognition approaches. It accomplishes this by combining acoustic phonetics principles and ideas with pattern recognition technologies.

III. CONCLUSION

Based on the foregoing discussion, it is possible to conclude that speech recognition applications are becoming increasingly important and useful in today's world. Speech recognition is the act of transforming input signals (typically speech) into well-structured word sequences, according to the findings of this study. It is worth noting that these sequences are created in the form of algorithms. Essentially, these algorithms convert speech into words and vice versa, resulting in more coherent, accurate, and correct speech recognition. Speech recognition has been identified as one of the most difficult challenges, and several techniques and approaches have been developed to address this issue. Among all of these paradigms and models, artificial intelligence is regarded as one of the most dependable and appropriate approaches.

REFERENCES

[1]. Das, Sanjib. "Speech Recognition Technique : A Review." International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 3, May-Jun 2012, pp.2071-2087.

[2]. Shoba Sivapatham, Rajavel Ramadoss. "Performance improvement of monaural speech separation system using image analysis techniques" , IET Signal Processing, 2018.

[3]. Priyadarshani, Nirosha & Dias, N. G. J. & Punchihewa, A.. (2012). Genetic Algorithm approach for sinhala speech recognition. Midwest Symposium on Circuits and Systems. 896-899. 10.1109/MWSCAS.2012.6292165.

[4]. Haridas, Arul & Marimuthu, Ramalatha & Sivakumar, Vaazi. (2018). A critical review and analysis on techniques of speech recognition: The road ahead. International Journal of Knowledge-based and Intelligent Engineering Systems. 22. 39-57. 10.3233/KES-180374.

[5]. Desai, Nidhi, Kinnal Dhameliya, and Vijayendra Desai. "Feature extraction and classification techniques for speech recognition: A review." International Journal of Emerging Technology and Advanced Engineering 3.12 (2013): 367-371.