Available online at http://www.advancedscientificjournal.com
http://www.krishmapublication.com
IJMASRI, Vol. 2, issue 5, pp. 608-612, May -2022
10.53633/ijmasri.2022.2.5.013

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY ADVANCED SCIENTIFIC RESEARCH AND INNOVATION (IJMASRI)

ISSN: 2582-9130 **IBI IMPACT FACTOR 1.5** **DOI: 10.53633/IJMASRI**

RESEARCH ARTICLE

VOICE RECOGNITION SYSTEM

Santosh Kumar Srivastava[1], Mohd. Sahil[2], Saif Ali[3], Sangeet Sachdeva[4], Sumit Kumar[5]

[1]Assistant Professor, Department of CSE, Greater Noida Institute of Technology,
[2,3,4,5] UG Scholar Department of CSE, Greater Noida Institute of Technology

Abstract

Voice recognition is a technology used to recognize a particular voice. Voice signals form the basis for the speaker identification. We can use voice targeting in many application areas such as phone banking, phone shopping, database access and voice mail. One of the most powerful applications of voice recognition for security is where One can enter their voice for verification. Speech is the basic form of communication between people. Speech recognition is the process of converting speech sounds into corresponding text. Speech recognition technology has grown dramatically over the last few years. However, there are many important research challenges e.g. differences in speaker and language, environmental sound and word size etc. The purpose of this paper is to present a holistic view of the acceptance of speech that describes various processes and summarizes the various methods used in the standard speech system.

Keywords: Speech recognition, modeling, speech processing, training and assessment.

Introduction

Speech recognition, popularly also known as Automatic Speech Recognition (ASR) is the process of converting speech signal to a sequence of words by means of an algorithm implemented as a computer program. Speech processing is one of the major fields of signal processing. Speech recognition area aims at to develop techniques for speech input to a machine (Gaikward 2010). The early computer systems were limited in scope and power. But the revolution in computer technology has evolved the field of automatic speech recognition. Now a day's it's easy to store huge database for speech recognition due to advancement in computer technology. Language is basic medium for communication, so it's better to expect human computer interfaces in native languages(Samudravijaya 2016). There are only few languages on which speech recognition systems have been developed. So a lot of scope is there to build speech recognizers in native languages (Naziya and Deshmukh 2016). Advancement in statistical modeling of speech has gained a widespread application in the field of speech recognition. Automatic speech recognition has reduced the human efforts in many fields, such as automatic call processing in telephone networks, data entry, voice dictation; query based updated travel information and reservation, natural language understanding and translators etc. Speech recognition technology has also an extensive use in telephone networks to automate and enhance the operator services. (Campbell and Sturim 2006; Brady and Brandstein 2008; Anusuya and Katti 2009). This paper highlights the basic building blocks of speech recognition systems, technological progression and problems in path ofautomatic speech recognition.

Literature Review

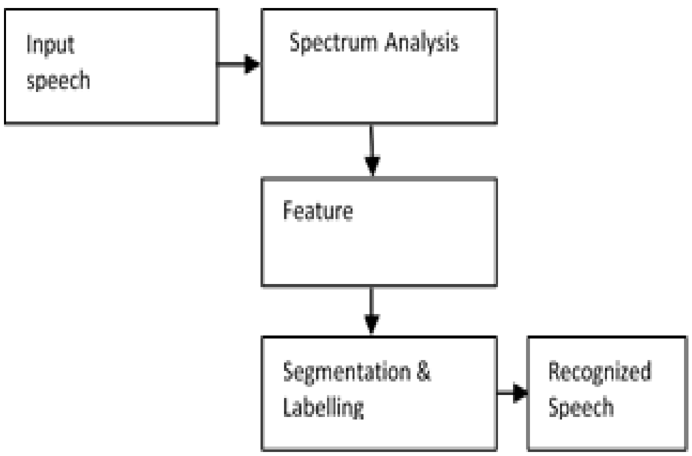
(Thiang, *et al.*, 2011) presented speech recognition using Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) for controlling movement of mobile robot. Input signals were sampled directly from the microphone and then the extraction was done by LPC and ANN. Ms.Vimala.C and Dr.V.Radha (2012) proposed speaker independent isolated speech recognition system for Tamil language. Feature extraction, acoustic model, pronunciation dictionary and language model were implemented using HMM which produced 88% of accuracy in 2500 words [29]. Cini Kurian and Kannan Balakrishnan (2012) found development and evaluation of different acoustic models for Malayalam continuous speech recognition. In this paper HMM is used to compare and evaluate the Context Dependent (CD), Context Independent (CI) models and Context Dependent tied (CD tied) models from this CI model 21%. The database consists of 21 speakers including 10 males and 11 females [7]. Suma Swamy et al. (2013) introduced an efficient speech recognition system which was experimented with Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), HMM which recognize the speech by 98% accuracy. The database consists of five words spoken by 4 speakers at ten times. (Annu Choudhary *et al.*,2013) proposed an automatic speech recognition system for isolated and connected words of Hindi language by using Hidden Markov Model Toolkit (HTK). Hindi words are used for dataset extracted by MFCC and the recognition system achieved 95% accuracy in isolated words and 90% in connected words (Naziya and Deshmukh 2016). Preeti Saini *et al.*, (2013) proposed Hindi automatic speech recognition using HTK. Isolated words are used to recognize the speech with 10 states in HMM topology which produced 96.61%. (Md. Akkas Ali *et al.*, 2013) presented automatic speech recognition technique for Bangla words. Feature extraction was done by, Linear Predictive Coding (LPC) and Gaussian Mixture Model (GMM). Totally 100 words recorded in 1000 times which gave 84% accuracy. (Maya Moneykumar, *et al.*, 2014) developed Malayalam word identification for speech recognition system. The proposed work was done with syllable based segmentation using HMM on MFCC for feature extractio. Jitendra Singh Pokhariya and Dr. Sanjay Mathur (2014) introduced Sanskrit speech recognition using HTK. MFCC and two state of HMM were used for extraction which produces 95.2% to 97.2% accuracy respectively [16]. In 2014, Geeta Nijhawan et al. developed real time speaker recognition system for Hindi words. Feature extraction done with MFCC using Quantization Linde, Buzo and Gray (VQLBG) algorithm. Voice Activity Detector (VAC) was proposed to remove the silence.

Methodologies

VRS methodologies can be usually classified in to three categories namely, Acoustic– Phonetic approach, pattern recognition approach and Artificial intelligence approach.

A. Acoustic-Phonetic Approach

Acoustic Phonetic approach is a rule based approach of speech recognition. According to this approach there exist finite, distinctive phonetic units in spoken speech utterances. These acoustic properties are present in speech signal. Acoustic – Phonetic approach involves the spectral analysis of speech signal to extract set of features for segmentation and labeling of speech signal into stable acoustic regions. In this way a valid word from segmentation to labeling is produced.

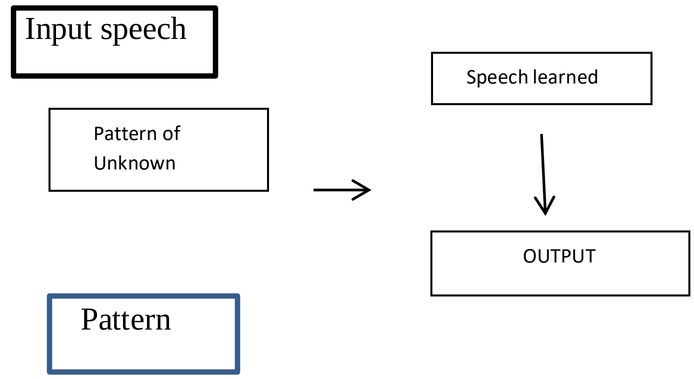


B. Pattern recognition Approach

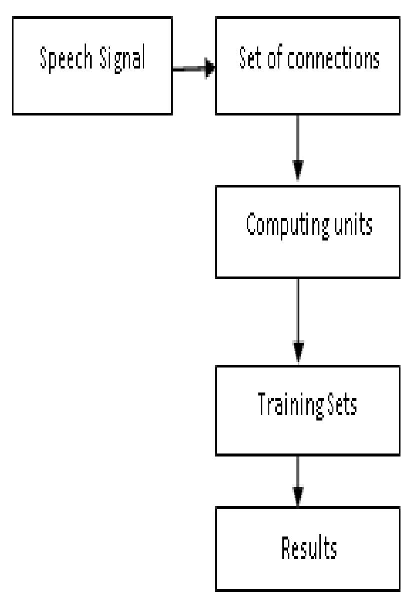
This approach is completed in two steps. First step is used to train speech recognizers with each possible pattern. The second stage provides a direct comparison between unknown speeches to the pattern learned in training stage. This approach have well defined mathematical framework which maintain consistent speech pattern to be recognized. The pattern recognition can have any of the form, either a speech template or a statistical modeling.

C. Artificial Intelligence Approach

It is a type of hybrid approach that exploits the idea of Acoustic phonetic approach and pattern recognition. Generally pattern matching is based on dynamic time warping (DTW) and hidden markov models.



In DTW speech recognition works with classes. Each class can be represented by one or more templates. As the number of templates increases it improes the system modelling. In advanced systemhidden marko model (HMM) is preferred over DTW, due to improved and lower memory requiremens. This approach is used for complex tasks but it is not so efficient when data set are large. Phoneme recognition is basic approach of artificial neural networks. This is done by technique of intelligence, analysing and visualizing the input speech. The network contains a large number of neurons. Each neurons counts for nonlinear weight of inputs and then send result to outgoing units. Training sets obtained, this way helps to assign values to input and output neurons.



A. *Deep Learning (HMM-DNN)*

Deep learning or unsupervised feature learning is new to the area of machine learning. It is the latest technology for speech recognition that has replaced the older one in all sense. It has generally three types of deep architectures, generative, discriminative and hybrid. The first type is used to provide high order correlation properties of data. The second type is intended to provide discriminative power for pattern classification and characterizing the posterior distributions of class labels. In third type goal is achieved through the composite output of first and second deep architectures.

Benefits of Voice Recognition System

1. Voice recognition technology is faster: Speaking isnormally faster than writing or typing among most of theindividuals – speech recognition software offers to getwords into documents without delay.
2. Accuracy is fairly good: Although transcripts from automatic speech recognition software need to be proofed and checked for quality, its quality is fairly good.
3. Hands-free, focused work: Voice recognition makes it possible for the dictating professional to focus on his or ercore function, without the need for paying attention to theroutine task of typing – you just need to dictate while evenattending to your core activities.
4. Spelling: Speech to text process relieves individuals fromhaving to pay attention to spelling – being able to dictatedirectly into the digital device ensures that spelling errorsare reduced considerably.

Conclusion

In this paper the basics are discussed and their recent progress is investigated. Different ways available by building a voice recognition system based on a modified feature translation process and speech recognition language system are compared in this paper. Voice recognition is a computer analysis of the human voice, mainly aimed at translating words as well phrases and regular identification of the speaker on the basis of the individual details included the waves of speech. This process makes it possible to use the voice of the presenter and is easy to verify personality. It provides access to control over various services such as google voice, ecommerce, window talk recognition, m-commerce, automation, home automation and safety management etc. This paper provides Review of various word recognition systems and speakers. Speech is a basic form of human interaction creatures, so it's a very user-friendly interface. Although the sector has gained a lot permission to modify apps and applications but there are a few parameters that affect accuracy as well the effectiveness of the speech recognition system. The most varied of speech involves the level of speech, of nature situations, channel and context of speech. The robustness of the speech system is subject to certain / Speech signal features. To improve the power of the speech recognition system, it is necessary to design the speech they see in local languages. Multilingualism is a revolutionary new field in the field of speech recognition. There are many development and research in the field of foreign languages but to improve its power and usefulness to indigenous peoples, it is important to use this technology in the indigenous languages. This paper offers various voices and speaker reviews recognition programs.

References

1. S. K. Gaikward(2010). “A review on speech recognition technique”, International journal of Computer Applications, vol. 10, no. 3, November 2010.
2. Samudravijaya, K (2016). “Speech and Speaker recognition tutorial” TIFR Mumbai 4000005.
3. Naziya, S. and Deshmukh, R. R.(2016). “ Speech Recognition System- A Review”, IOSR Journal of Computer Engineering (IOSR-JCE), vol. 18,issue 4, (Jul.-Aug. 16), pp. 01-09.
4. Campbell,W.M and Sturim, D.E(2006). “The MIT- LL/IBM speaker recognition system using high performance reduced complexity

5. recognition" MIT Lincoln Laboratory IBM 2006.
6. Brady, K and Brandstein, M.(2008) "An evaluation of audio-visual person recognition on the XM2VTS corpus using the Lausanne protocol", MIT Lincoln Laboratory, 244 Wood St., Lexington MA.
7. Anusuya, M. A and Katti,S.K. (2009). " Speech Recognition by Machine : A Review", International journal of Computer Science and Information Security, vol. 6, no. 3, 2009 , pp.181-205.
8. Shweta, D and Rajni, M (2014). "Speech Recognition Techniques: A Review", International journal of advanced research in computer science and software engineering, vol. 4, issue 8.
9. Kuldeep, K and Aggarwal, R. K.(2012). "A Hindi speech recognition system for connected words using HTK", Int. J. Computational systems Engineering, vol. 1, No. 1. , Haryana, India, 2012.
10. Mayank, D and Aggarwal, R. K.(2008). "Implementing a speech recognition system interface for Indian languages", proc.of the JCNLP-08 workshop on NLP", , Hyderabad, India, January 2008, pp. 105-112.
11. H. Hermansky (1990). "Perceptually predictive analysis of speech.", Journal of Acoustic Society of America, vol. 87, 1990, pp. 1738-1752.
12. Santosh Kumar Srivastava,(2020). "S-Transform and Gaussian Mixture Model for Acoustic Scene Classification", published in International Journal of Advances in Signal and Image Sciences, **e-ISSN: 2457-0370,** Vol.6, No.1.
13. Santosh Kumar Srivastava, Dr. Yogesh Kumar Sharma, Dr. Sheo Kumar, (2020). "Using Of WEKA Tool In Machine Learning: A Review" Published in International Journal of Advanced Science and Technology Vol. 29, No. 6, (2020), pp. 8604-8614 (SCOPUS Indexed), ISSN: 2005-4238 IJAST.
14. Santosh Kumar Srivastava, Dr. Yogesh Kumar Sharma, Dr. Sheo Kumar,(2019). "Precision enhancement of Intrusion detection system through outlier detection and feature classification" Published in International Journal of Control and Automation (SCOPUS Indexed) Vol. 12, No. 6, (2019), pp. 820-830, ISSN: 2005-4297 IJCA

*****