A Study of Speech Recognition Techniques

Dr. R. Vishnupriya[1], Dr. J. Viji Gripsy[2], Dr. S. Karpagavalli [3]
[1, 2, 3] Dept of Computer Science
[1, 2, 3] PSGR Krishnammal College for Women, Coimbatore, India

Abstract- *This paper presents the basic idea of speech recognition, types of speech recognition, prons and cons with various pattern matching approaches for recognizing the speech of the different speaker. It also discuss about the process of speech recognition, types of speech recognition and speech databases. Now a days research in speech recognition is motivated for ASR system that supports speaker independent operations and continuous speech in different language.*

Keywords- Speech Recognition, Types, approaches and databases.

I. INTRODUCTION

Speech Recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine – readable format. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. More sophisticated software has the ability to accept the natural speech.

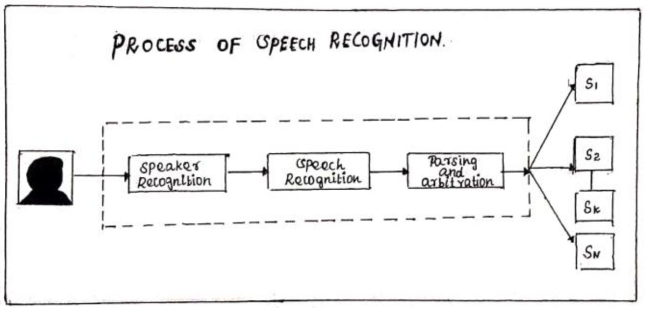
The first speech recognition systems could understand only digits. Bell laboratories designed the "AUDREY" system in 1952, which recognized digits spoken by a single voice. Ten years later, IBM demonstrated the World's Fair its "SHOEBOX" at 1962. This system will understood 16 words spoken in English. Speech Recognition is also known as Automatic Speech Recognition (ASR), Computer Speech Recognition or Speech to Text (STT). Speech recognition applications include Voice User Interface such as voice dialing (eg ., "call home").

II. NEEDS OF SPEECH RECOGNITION

Speech Recognition is software that allows the user to interact with their mobile devices through speech. It enables a machine to single our words or phrases in a spoken language, after that it converts into machine readable format. As an inter-disciplinary sub-field of computational linguistics, it develops technologies that first recognizes then converts spoken language into text by computers. Speech Recognition system had found a range of use in computerized games and toys, Control of Different Instruments, data collection and dictation. These features are also proved to be much need among those who could not obtain keypads and among with certain disabilities. Speech Recognition is the process by converting the human speech as understandable for machines as signals then to text. Speech Recognition has a wide range of use and it deployed in contact centers, IVR systems, mobile and embedded devices, dictation solutions and assistive applications.

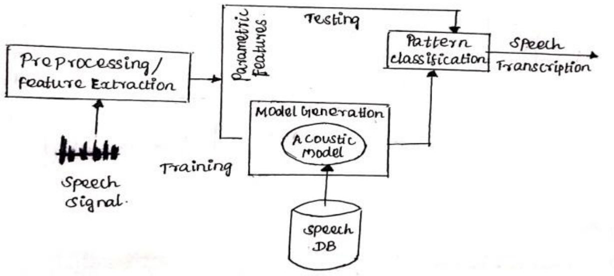
III. SPEECH RECOGNITION PROCESS

Speech processing technologies are used for digital speech coding, spoken language dialog systems, text-to-speech synthesis, and automatic speech recognition. Speech processing is the study of speech signals and processing methods. Speech Recognition signals are usually processed in a digital representation, so speech processing can be regarded as digital Signal Processing [10]. Speech processing includes the acquisition, manipulation, storage, transfer and output of speech signals. Speech processing is a "NATURAL" way to interact with a computer. Speech is hands-free, eyes-free, fast and intuitive. Speech processing would be much easier if there were linear relationships between articulations and acoustics, and between acoustics and perception. This would greatly facilitate automatic speech synthesis and recognition respectively. Speech processing and Natural Language processing (NLP) allow intelligent devices like smart phones to interact with users via verbal language. The most well-known example of speech recognition technology on a mobile device Apple's voice recognition service.



Process of Speech Recognition(fig., 1.1)

Most of the speech recognition systems consist of three major modules like Signal Processing front-end, acoustic modeling and language modeling. Speech recognition is the process by which computers maps an acoustic speech signal to some form of abstract meaning of the speech. This process is highly difficult. Since sound has to be matched with stored sound bites with further analysis has to be done because sound bites do not match with the pre-existing sound pieces. There are two main phases in a speech recognition system. Training phase and testing phase for speech recognition. During the training phase, first the features are extracted from the all speech signals using various feature extraction techniques such as MFCC, LPC, LDA and RASTA. These features are in the form of vector. In the training the vector is generated from the speech signal of each word spoken by the user. The training vector which distinguishes different words based on its class. This training vector is used in the next phase of recognition. During the speech recognition phase, the user speaks any word for the system which was trained. A test pattern is generated for that word. This test pattern extracts the feature of the word used for testing. In this way the test pattern is tested against the training vector by using various classifications such as SVM, KNN, HMM and ANN [4]. If the testing word pattern matches with the training pattern class and it means the particular pattern is recognized from training phase and the corresponding pattern is recognized from training phase and that corresponding pattern is displayed as the output. The system block diagram of speech recognition process is show in Fig. 1.2



Block diagram of speech recognition process. (Fig.1.2)

IV. TYPES OF SPEECH RECOGNITION

There are two types of speech recognition.

- Speaker - dependent.(It is commonly used for dictation)
- Speaker – independent. (It is commonly used in telephone applications)
- Discrete speech recognition
- Continuous speech recognition
- Natural language

A. Speaker - Dependent

This software works by learning the unique characteristics of a single person's voice, in the similar way of voice recognition. New users must "train" the software by speaking to it, so the computer can analyze easily how the person speaks [8]. This means the user have to read a few pages of the text to the computer before they use the speech recognition software. This model recognizes speech patterns form only one persons. But both models use mathematical and statistical formulas to yield the work match for speech. A third variation of speaker models is now emerging, called SPEAKER ADAPTIVE [1]. The simple way to say about speaker – dependent software recognized most users voices with training. Generally speaker dependent are more accurate for the correct speaker, but less accurate for other speakers. They assume speaker will speak in a consistent voice and tempo. This often requires a user reads a series of words and phrases so the computer can understand the users voice.

B. Speaker – Independent

This software is designed to recognize anyone's voice, so no training is involved. It is the only real option for applications such as interactive voice response systems. In this model businesses can't ask callers to read the pages of the text before using the system. The speaker – independent software is generally less accurate than speaker – dependent. Speech Recognition engines are the speaker independent generally deal with the fact by limiting the grammars. By using a smaller list of recognized words, this speech engine is more likely to correctly recognize what a speaker said. This software deals with most of the IVR systems, and an application where a large number of people will be using the same system. The speaker-independent usually begin with a speaker independent model and adjust these models more closely to each individual during a brief training period. The simple way to say about the speaker-independent software recognizes most users voices without training [9].

C. Discrete Speech Recognition

In Discrete speech recognition the user must pause between each word so that the speech recognition can identify each word separately.

D. Continuous Speech Recognition

This continuous speech recognition can understand a normal rate of speaking. Continuous speech recognition allows users to speak almost naturally. The recognizer with continues speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries.

E. Natural Language

In natural language the speech recognition not only understand the voice but also retrun answers to questions or other queries that are being asked.

V. APPLICATIONS OF SPEECH RECOGNITION

Speech Recognition technology and the use of digital assistants have moved quickly from our cell phones to our homes, and its application in industries such as business, banking, marketing, and healthcare is quickly becoming apparent. Speech recognition applications include

- Voice dialing(eg., “Call home”)
- Call routing(eg., “I would like to make a collect call”)
- Simple data entry(eg., entering a credit card number)
- Preparation of structured documents(eg., A radiology report)
- Speech-to-text processing(eg., word processors or emails), and
- In aircraft cockpits (usually termed Directed Voice Input)

Speech recognition technology in the workspace has evolved into incorporating simple tasks to increase efficiency, as well as beyond tasks that have traditionally needed humans, to be performed.

Some of the examples of the workplace of digital assistants are.

- Search for reports or documents on your computer.
- Create a graph or tables using data
- Dictate the information you want to be incorporated into a document.
- Print documents on request.
- Start video conferences.
- Schedule meetings.
- Record minutes.
- Make travel arrangements

Banking

The aim of banking and financial industry in the speech recognition applications to reduce friction for the customer. Voice activated by banking could largely reduce the need for human customer service and lower employee costs. A personalized customer satisfaction and loyalty.

How speech recognition could improve banking:

- Request information regarding your balance, transactions and spending habits without having to open your cell phone.
- Make payments
- Receive information about your transaction history [2]

Marketing

In marketing Voice search has the potential to add a new dimension to the way marketers to reach their consumers. If the change in how people are going to be interacting with their devices, marketers should look for developing the trends in user data and behavior[5].

Data – with the speech recognition, there is a new type of data available for marketers to analyse. Speech patterns and vocabulary can be used to interpret a consumers location, age and other information regarding their demographics such as their cultural affiliation.

Behaviour – While typing necessitates a certain extent of brevity, speaking allows for longer, and more conversational searches. In marketing it may need focus on long – tail keywords and producing conversational content to stay ahead of these words.

Some of the common applications are

Medical Transcription
Military
Telephony and other domains
Serving the disabled.

VI. APPROACHES TO SPEECH RECOGNITION BY MACHINE

Basically there are three approaches to speech recognition. They are

Acoustic Phonetic Approach
Pattern Recognition Approach
Artificial Intelligence Approach

A. Acoustic Phonetic Approach

The earliest approaches to speech recognition which was based on finding speech sounds and providing appropriate labels to these sound. This is the basis of the acoustic phonetic approach. Which postulates that there exist finite, distinctive phonetic units in spoken language and that these units are characterized by a set of acoustics properties that are manifested in the speech signal over time. The first step in the acoustic phonetic approach is a spectral analysis of the speech combined with the feature detection that converts the spectral measurements for the set of features that describe the broad acoustic properties of the different phonetic units. The next step is a segmentation and labeling phase in which the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels to each segmented region. This result in the phoneme lattice characterization of the speech. The last step in this approach attempts to determine a valid word from the phonetic label sequences produced by the segmentation to labeling. In this validation process, linguistic constraints on the task (i.e., the vocabulary, the syntax and other semantic rules) are invoked in order to access the lexicon for word decoding based on the phoneme lattice. The acoustic phonetic approach has not been widely used in most commercial applications [3].

B. Pattern Recognition Approach

The pattern recognition approach involves two essential steps namely, pattern training and pattern comparison. The essential feature of this approach is that is uses a well formulated mathematical framework and establishes consistent speech pattern representations, for reliable pattern comparison, from a set of labeled training samples via a formal training algorithm. A speech pattern representations can be in the form of speech template or a statistical mode (eg., A HIDDEN MARKOV MODEL or HMM) and it can be applied to a sound or a phrase. In the pattern of the comparison approach, a direct comparison is made between the unknown speeches with each possible pattern learned in the training stage in order to determine the identity of the unknown according to the goodness of march of the patterns. The pattern – matching approach has become the predominant method for speech recognition in the last six decay [8].

C. Artificial Intelligence Approach

The artificial Intelligence approach in 1997 is based on hybrid of the acoustic phonetic approach and pattern recognition approach. In this approach, it exploits the ideas and concepts of Acoustic phonetic and pattern recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Some speech researchers developed recognition system that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based approaches have been very effective in the design of a variety of speech recognition systems, they provides linguistic and phonetic literature provided sights and understanding the human speech processing [7]. However, this approach had only limited success, largely due to the difficulty in quantifying expert knowledge. Another difficult problem is the integration of many levels of human knowledge – phonetics, phonotactics, lexical access, syntax, semantics and pragmatics. This form of knowledge applications makes an important distinction between knowledge and algorithm-Algorithm enable us to solve the problems. Knowledge enable the algorithms to work better. This form of knowledge based system enhancement has contributed considerably to the design of all successful strategies reported [6].

VII. LITERATURE SURVEY ON SPEECH RECOGNITION

In this literature survey “Speech Recognition in the Electronic Health Record” whose authors are Sherry Doggett, Julie A. Dooling (RHIA), Susan Lucci (RHIT, CHPS, AHDI-F) have done work on “Speech Recognition in the Electronic Health Record(EHR)” using front-end speech recognition(FESR) and back-end speech recognition (BESR) technologies help in the production of legible and comprehensive documents. It serves as a productivity tool to help lower costs and increase productivity, especially when compared to the manual labor required by the traditional dictation and transcription in the field of healthcare.

The author of literature review on automatic speech recognition is WIQAS GHAL, have done work in the field of ‘Automatic Speech Recognition(ASR)’ for developing an effective ASR for different languages and to show technological perspective on ASR for different languages and to show technological perspective of ASR in different countries. They have used artificial neural networks (ANNs), mathematical models of the low-level circuits in the human brain, to improve speech-recognition performance, through the model known as the ANN-Hidden Markov Model (ANN-HMM) which has shown great improvements in large-vocabulary speech recognition systems.

VIII. TOOLS FOR SPEECH RECOGNITION

Some of the common software tool used to help you to work faster

- Google Dogs Voice Typing
- Dragon Professional Individual
- Braina Pro
- Speech notes
- e-speaking

- Voice Finger
- Apple Dictation
- Windows Speech Recognition

IX. CONCLUSION

This work deals with the various type of speech recognition, applications of speech recognition, approaches to speech recognition by machine, methods for speech recognition, literature survey on speech recognition, speech databases, performance of speech recognition and tools for speech recognition. We conclude that if the machine could successfully pretend to be human to a knowledge observer than you certainly should consider it intelligent. Speech recognition systems are an indispensable part of the ever-advancing field of human computer interaction. Needs greater research to tackle various challenges.

REFERENCES

[1] Malay Kumar , R K Aggarwal, Gaurav Leekha and Yogesh Kumar “Ensemble Feature Extraction Modules for Improved Hindi Speech Recognition System”, International Journal of Computer Science Issues, Vol.9, Issue 3,no 1, May 2012,

[2] Pukhraj P. Shrishrimal, Vishal B. Waghmare, Ratnadeep Deshmukh, “Indian Language Speech Database: A revuew”, 1 international Journal of Computer Application, vol 47-no.5,June 2012.

[3] http://en.wikipedia.org/wiki/Speech _ recognition

[4] https://www.scribd.com/doc/130376790/Speech-Recognition

[5] “speaker Independent Connected Speech Recognition-Fifth Generation Computer Corporation”.Fifthgen.com.

[6] http://www.speechrecognition.com

[7] https://www.google.co.in/?gfe_rd=cr&ei=GbHdU9f1MtKAoAoW64GADg&gws_rd=ssl

[8] Koester HH. User performance with speech recognition: a literature review. Assist. Technol. 2001;13(2):116-30.

[9] Abishek Thakur, Naveen Kumar, “Automatic Speech Recognition System for Hindi Utterance with Regional Indian Accents: A Review”, International Journal of Electronics & Communication Technology, Vol. 4, April –June 2013.

[10] Sadaoki Furui, November 2005, 50 years of Progress in speech and Speaker Recognition Research, ECTI Transaction on computer and Information Technology,vol,1.No.2.