Speech Recognition using Artificial Intelligence

Arjun Raja Y[1], Ashika N Bhandary[2], Anitha P[3]

Dept of ISE, JSSATE, Bangalore[1,2,3,4]

Assistant Professor, Dept. of ISE, JSSATE, Bengaluru[3]

Abstract: The invention the computer has proved to be the most significant invention of the twentieth century and humans have made a constant effort to improve the intelligence of the computer. This effort will continue unless they produce computers as intelligent as humans. This will be implemented by a familiar concept, which brings us to our topic of discussion: ARTIFICIAL INTELLIGENCE. Out of the various resourceful branches under artificial intelligence, this paper talks about speech recognition. It is the process of enabling a computer to recognize and revert to the sounds produced in human speech. Speech recognition helps the users to convert the voice provided by them to the computer into a word document. It's time efficient since it does not require much physical effort. The applications include controlling car systems, health care system, military and guiding the specially abled. The paper also talks about the two most popular speech recognition software's available for the masses now are Apple Corp's Siri and Google's virtual assistant Google Now. Finally, the paper concludes with how speech recognition will play a significant role in creating a technological environment that is comparatively faster and enduring.

I INTRODUCTION

The area of artificial intelligence is expanding exponentially. So are the resources being provided by them. As we know, natural means of communication used by humans to interact with each other are speech and gestures. Artificial intelligence includes an efficient and precise man machine interfacing (MMI). For a machine to be able to interpret human speech, speech recognition technology is required. Speech recognition technology has been under development for many years but it had not been established enoughto be used with PCs until recently.

To make speech interfaces practical for frequent use the two major factors involved are accuracy and speed. Working on these two aspects will take speech recognition to a whole new level, which will be an imperative measure considering thepace at which our everyday tasks are being digitalized.

WHAT IS SPEECH RECOGNITION?
Speech recognition is a technology that enables a machine to understand and interpret human voice and perform the required task. It is also known as automatic speech recognition or computer speech recognition.

WHERE CAN IT BE USED?

A Hands free AI Assistant
A very famous AI assistant is the voice recognition provided by Apple, Siri (abbreviation for Speech Interpretation and Recognition Interface). Apple's Siriis based around voice command—user says something and the software responds in kind. In a recent upgrade Siri has been provided the privilege of identifying and authentication of the device. Apple recently included voice recognition on Siri, this has opened up the possibility of voice-based unlock for the iPhone.

Google's Google now is an intelligent personal assistant developed by Google Corp. Unlike Siri which sources its information from specific websites, Google now sources its info from any website it deemed relevant by tracking the user activities.

Voice Recognition to Protect your Private Accounts
Speech recognition is usually coupled with face recognition to create a multi-factor system. The reason behind this is as you speak for authentication the face recognition software will see your mouth moving and make sure it's legitimate.

Serves as a guide to the Specially-Abled
It helps people with mobility and sensory impairments to access computer technologies, making them lead a better lifestyle.
 It also plays a major role of acceleration in fields like aerospace, military, Healthcare systems etc.
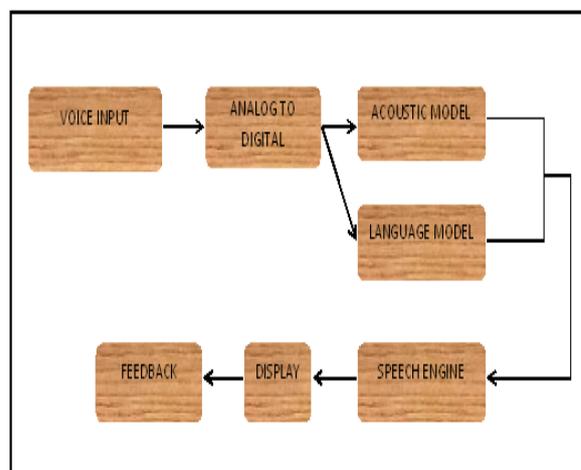
Speech recognition can be implemented in real-word through the two following techniques:

ACOUSTIC MODEL

An acoustic model can be established by collecting the voice recordings of speech along with it's textual interpretation and providing a statistical portrayal of all the adios that make up the words, using the software.

LANGUAGE MODEL

Language models are made use of in various natural language computing practices such as speech recognition record the attributes of a certain language and anticipate the upcoming word in the speech sequence pattern

TYPES OF SPEECH RECOGNITION:

There are two types of speech recognition:

SPEAKER DEPENDENT

It's generally used for dictation software's.

The software used in this practice by capturing the unique properties of an individual.

Users new to the software must first coach the software by speaking to it constantly so the machine can examine and get used to the way that person talks.

This can be done by making the user read a piece of content to the computer before they start using that software.

SPEAKER INDEPENDENT

The software can be used to identify anybody's audio, so no prior training is required for the software.

This proves to be a great help as the corporate level where demands like piece of content being read to the software will not be entertained.

The disadvantage of this software of the speaker dependent software is that it's very much less accurate comparatively.

So how the speaker independent drivers deal with this limitations by restricting the amount of grammar used. By limiting the number of words used to a list of engine recognized words, there's a higher probability that the driver identifies what the speaker has said appropriately.

These are mostly used in telephone applications.

DIFFERENT PROCEESSES INVOLVED IN SPEECH RECOGNITION:

DIGITIZATION

Analog to digital conversion

Sampling and quantizing: The mechanism of getting the domain digitized is called sampling.

The mechanism of getting the range digitized is called quantization.

SIGNAL PROCESSING

Eradicating unwanted noise and procuring the required speech.

PHONETICS

Instability in human voice.

PHONOLOGY

Comparing the contrast in distinctions of sounds.

With the use of sound, the meaning is encoded into any human language.

PRAGMATIC AND SEMANTICS

Pragmatics is the process of linking what the speaker is saying to what the sentence actually means.

Semantics deals with the meaning of words.

LEXICOLOGY AND SYNTAX

Lexicology is the analysis of the behavior, form and meaning of words. Syntax is how well the word can be arranged to get a sentence in any language.

STATISTICAL MODELLING SYSTEMS

Mathematical functions and probability are used by the system to deduce the most plausible outcomes.

The two governing models today are the Hidden Markov Model and the Neural Networks.

Both of these consist of complicated mathematical functions, but mostly they work with such information that is known to the system to get hold of the data unknown to it.

NUERAL NETWORKS

A branch of statistical models will be called as neural if they involve a collection of adaptive weights, that is, numerical parameters tuned by learning algorithm and are able ofrounding off the nonlinear functions into their inputs.

The adaptive weights that we talked about are abstractly link stability between neurons, that get activated while the training and prediction is inprocess.
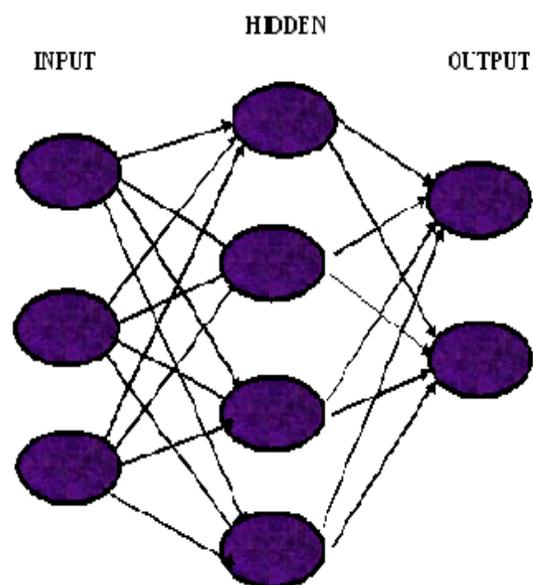

HIDDEN MARKOV MODEL

Here, every phoneme is a like a boogie of a train and the complete train is the word.

This chain splits into various branches as the process of trying to match the digital sound to that of the phoneme that is likely to occur next.

While the process is going on, depending on the built-in dictionary and user coaching, a probability score is assigned to every phoneme.
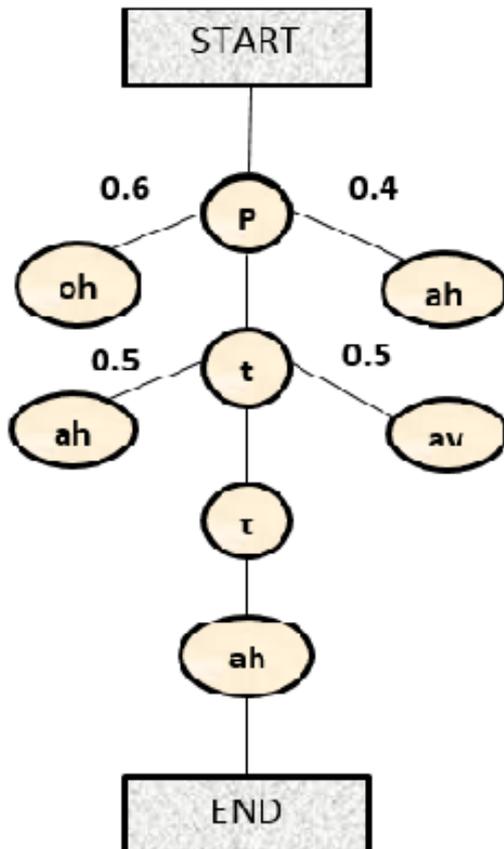
PROGRAM TRAINING

The process proves to be more complex while working on phrases or sentences: because the system must itself figure out where the previous word stops and the next one starts. The statistical working requires a lot of ideal training data to meet its optimal peak.

Occasionally on the order of a huge number of hours of human interpreted audio and a huge chunk of text data. Such training data will be made use of to form word lists or multi-word probability networks and acoustic model of words.

All these details will help you differentiate between a system that is performing well and a system with a poor performance even though they might be using the same algorithm at a given instance

Weaknesses and flaws
Low signal-to-noise ratio
The program should be able to listen the words being said distinctly and any unrequired noise will just act as intrusion. And here the required signal received compared to the noise intrusion is pretty low as they aren't able to reduce it yet.

Overlapping speech
The systems being used correctly aren't advanced enough to differentiate between two set of speeches provided to it simultaneously by multiple users. It is poor at handling the overlapping.

Intensive use of computer power
People are becoming too much technology dependent which might prove to be harmful in the longer run. Since speech recognition doesn't need much physical efforts, its making human lazier and more prone to having an unhealthy lifestyle.

Homonyms
Words that sound similar but spell and mean different pose a major problem since the system are developed enough to differentiate between them distinctly.
For example: The words like "their" and "there", "air" and "heir", "be" and "be" etc.

MAJOR CHALLENGES
One of the most primitive and most difficult challenges will be creating a system that can accurately distinguish and recognize slang, dialects, accents and background noise, which is a major roadblock in the real world.
The different grammatical structures used by languages can also pose a problem. For example, Arabic sometimes uses single words to convey ideas that are entire sentences in English.

There are Hundreds of thousands of languages and dialects all over the world; to make speech recognition auser phenomenon it is important to be able to reach the masswhich is a pretty tall order in the current scenario.

The Future of Speech Recognition
In this day and age where the whole world is constantly connected it is crucial to have effective communication between persons, whether or not they share a similar language, to overcome this linguistic barrier a speech recognizing tool which can translate speech at around ninety percentage of accuracy is required.

The United States' Defense Advanced Research Projects Agency (DARPA) has three teams of researchers working on a program (Global Autonomous Language Exploitation (GALE)) that will take in streams of information from foreign news broadcasts and newspapers and translate them.

DARPA is also funding an R&D effort called TRANSTAC to enable the soldiers to communicate more effectively with civilian populations in non-English-speaking countries.

CONCLUSION

The agenda behind Artificial intelligence is for the computer to understand the meaning behind a command and to act accordingly without the need for human inputs. Machines already see better than humans, recognize objects better, and can listen and hear better. Eventually they will also understand meaning better.

Voice recognition in machines is getting very good and is going to get so good that it will completely change the way humans interact with their computing devices.

For these objectives to be completed is a huge task ahead of us in terms of computational power and software sophistication, but with the rapid growth that we see that speech recognition development is undergoing today it is not a long wayfrom achieving true artificial intelligence.

REFERENCES

[1] http://en.wikipedia.org/wiki/Hidden_Markov_mode
[2]   http://electronics.howstuffworks.com/gadgets/high-tech-gadgets/speechrecognition.html
[3] https://www.slideshare.net/asertseminar/speech-recognition-technology
[4] https://www.google.co.in/search?q=Defense+Advanced+Research+Projects+Agency&oq=Defense+Advanced+Research+Projects+Agency&aqs=chrome..69i57j0l5.1307j0j8&sourceid=chrome&ie=UTF-8