Recog- A Speech Recognition Artificial Intelligent Software

N. Sandhya
Assistant Professor,
Department Of Computer Science,
St. Joseph's College (Autonomous),
Langford Road, Shanthinagar,
Bangalore – 560027, India.

B. Nithya
Assistant Professor,
Department Of Computer Science,
St. Joseph's College (Autonomous),
Langford Road, Shanthinagar,
Bangalore – 560027, India.

Prasad. C. N
Assistant Professor,
Department Of Computer Science,
St. Joseph's College (Autonomous),
Langford Road, Shanthinagar,
Bangalore – 560027, India.

***Abstract*: This paper deals with a software called recog that was developed as a part of my artificial intelligence project. It is based on the concept of speech recognition. It recogonizes few words that are present in the program and perform a certain action. It is built on the visual studio platform. This paper also explains what is speech recognition? , the difference between speech recognition and voice recognition, why is speech recognition used, The artificial intelligence technique used in this project, Why this technique was chosen?, What are the alternative techniques that could have been used?.**

I. INTRODUCTION:

Speech recognition (SR) is the inter-disciplinary sub-field of computational linguistics which incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields to develop methodologies and technologies that enables the recognition and translation of spoken language into text by computers and computerized devices such as those categorized as Smart Technologies and robotics. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

•

Some SR systems use "training" (also called "enrollment") where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker independent"[1] systems. Systems that use training are called "speaker dependent".

Speech recognition applications include voice user interfaces such as voice dialing (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domotic appliance control, search (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

The term *voice recognition*[2][3][4] or *speaker identification* refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process. From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the world-wide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems. These speech industry players include Microsoft, Google, IBM, Baidu (China), Apple, Amazon, Nuance, IflyTek (China), many of which have publicized the core technology in their speech recognition systems being based on deep learning.

I What is Speech Recognition?

- 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.
- You can use your voice to control your computer. You can say commands that the computer will respond to, and you can dictate text to the computer.

You'll need to connect a microphone to your computer. Once you've got the microphone set up, you can train your computer to better understand you by creating a voice profile that your computer uses to recognize your voice and spoken commands. For information about setting up your microphone. After you've got your microphone and voice profile setup, you can use Speech Recognition to do the following:

- Control your computer. Speech Recognition listens and responds to your spoken commands. You can use Speech Recognition to run programs and interact with Windows. For more information about the commands you can use with Speech Recognition, see Common commands in Speech Recognition.
- Dictate and edit text. You can use Speech Recognition to dictate words into word-processing programs or to fill out online forms in a web browser. You can also use Speech recognition to edit text on your

computer. For more information about dictating text, see Dictate text using Speech Recognition.

Difference between voice recognition and speech recognition.

VOICE RECOGNITION	SPEECH RECOGNITION
Voice recognition typically disregards the language and meaning to detect the physical Person behind the speech.	Speech recognition strips out the personal differences to Detect the words.
voice recognition is independent Of language.	Speech recognition is language Dependent.

Why do you need a speech recognition software?

- Number of applications for users with and without disabilities.
- Speech-to-text has been used to help struggling writers boost their writing production
- And to provide alternate access to a computer for individuals with physical impairments
- Other applications include speech recognition for foreign language learning,
- Voice activated products for the blind,
- And many familiar mainstream technologies.
- Automated phone menus and directories,
- Voice activated dialing on our cell phones, and
- Integrated voice commands on Smartphones are just a few examples
- Medical and law professionals use voice recognition every day to dictate notes and Transcribe important information.
- Newer uses of the technology include military applications, navigation systems, automotive Speech recognition (Ford SYNC), 'smart' homes designed with voice command devices, and Video games such as End War, which allows the player to give orders to their troops using only their voice.

Advantages of Speech Recognition and Disadvantages Speech Recognition

- Reduces the time complexity
- Reduces the overhead on the user.
- Simple and interactive interface.
- Reduces the space Complexity.
- Used to search a remote Application or file.
- User with less GRAMMAR Knowledge can access the files and applications.
- User must follow the voice instructions specified in the software, else it leads to a Incorrect output or error.

II RECOG:

Applications Already present

- Amazon
- Gmail
- Facebook
- Flipkart
- Vlc
- Open run
- Chrome
- Excel
- Powerpoint word
- Twitter
- Recog

Applications embedded in forms

Recog:

- Web cam
- Media player
- Photo viewer
- Movie player
- Maps
- Translator
- Google search
- YouTube search
- Weather

Features

- It opens up any of the listed applications without the usage of the mouse or keyboard.
- The mouse is only used to enable the speech recognition or disable only.
- Whatever is spoken by the user is recognized by the speech recognition Engine and synthesizer.
- Then the text is compared with each of the switch case conditions and then the corresponding block is executed that is opening the required software or performing an operation.
- The close command closes the current form.

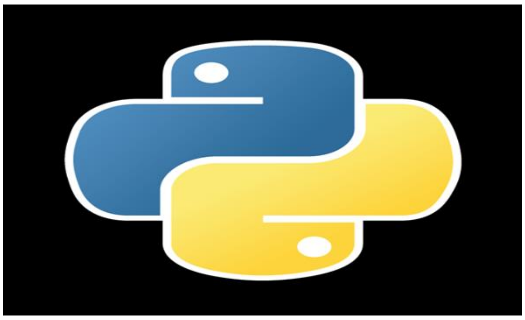
Goal of the project

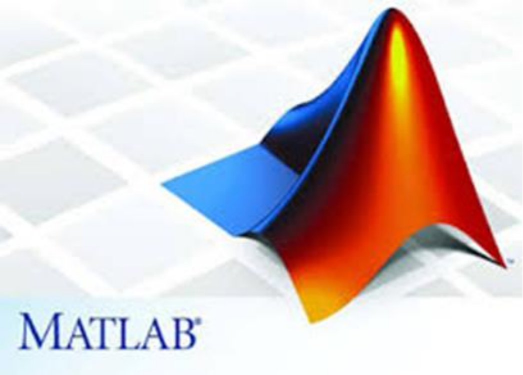
- Open and access applications without the help of a mouse or keyboard.
- Click pictures without having to hit the shutter button.
- Play a movie without having to browse or have the mouse click event happen.
- Play a song, view a picture, watch a movie by just speaking to the computer without having to search for it, or by clicking on the icon.

Apple's Siri, Google's now, and Microsoft's Cortana are smartphone "personal assistant apps"

Platform on which the project is built

Other platforms on which it can be built





Limitations

- You need to have the visual studio software in order to run this application. It is not an independent standalone software.
- To start the process of speech recognition the mouse is initially use- to enable speech recognition engine and synthesizer and also to disable.
- Not all the applications are embedded. Since Google api’s were required.

Applications

- In-car systems
- Health care-Medical documentation
- Military-High-performance fighter aircraft-control of communication radios, setting of navigation systems, and control of an automated target handover System.
- Telephony and other domains- SR in the field of telephony is now commonplace and in the field of computer gaming and simulation is becoming more widespread. Despite the high level of integration withword processing in general personal computing. However, ASR in the field of document production has not seen the expected [by whom?] increases in use.

III. TECHNIQUE USED FOR SOLVING THE PROBLEM: TEMPLATE MATCHING

- Template matching is the simplest technique and has the highest accuracy when used properly, but it also suffers from the most limitations.
- The first step is for the user to speak a word or phrase into a microphone.
- The electrical signal from the microphone is digitized by an "analog-to-digital (A/D) Converter", and is stored in memory.
- To determine the "meaning" of this voice input, the computer attempts to match the Input with a digitized voice sample, or template that has a known meaning.
- This technique is a close analogy to the traditional command inputs from a keyboard.
- The program contains the input template, and attempts to match this template with the
  Actual input using a simple conditional statement.
- This type of system is known as "speaker dependent." and recognition accuracy can be
  About 98 percent.

IV WHY THIS TECHNIQUE?

- It was easy to implement as a program
- Every word has a definite action to perform
- Only a confined set of words are chosen and those are given a particular action to be performed.
- Since many forms are present the search becomes easier using this technique.
- In each form the keywords used to perform an operation is different and the number of operations vary per form.
- This method is chosen in order to reduce the time taken for the existing system to open a file, retrieve, weather report etc.….
- This proposed system mainly aims at reducing the time taken to perform certain operations.

- The proposed software requires less memory space when compared to the existing software and more over it opens a video, audio file, retrieves weather report, opens certain websites like google, YouTube, twitter etc. with just voice commands.

V WHAT ARE THE ALTERNATE TECHNIQUES? FEATURE ANALYSIS:

This technique usually leads to "speaker-independent" speech recognition.

- Instead of trying to find an exact or near-exact match between the actual voice input and a previously stored voice template, this method first processes the voice input using "Fourier transforms" or "linear predictive coding (LPC)", then attempts to find characteristic similarities between the expected inputs and the actual digitized voice input.
- These similarities will be present for a wide range of speakers, and so the system need not be trained by each new user.
- The types of speech differences that the speaker-independent method can deal with, but which pattern matching would fail to handle, include accents, and varying speed of delivery, pitch, volume, and inflection.
- Speaker-independent speech recognition has proven to be very difficult, with some of the greatest hurdles being the variety of accents and inflections used by speakers of different Nationalities.

Simple pattern matching

- You'll have encountered it if you've ever phoned an automated call center and been answered by a computerized switchboard.
- Utility companies often have systems like this that you can use to leave meter readings, and banks sometimes use them to automate basic services like balance inquiries, statement orders, checkbook requests, and so on.
- You simply dial a number, wait for a recorded voice to answer, then either key in or speak your account number before pressing more keys (or speaking again) to select what you want to do.
- Crucially, all you ever get to do is choose one option from a very short list, so the computer at the other end never has to do anything as complex as parsing a sentence (splitting a string of spoken sound into separate words and figuring out their structure), much less trying to understand

Statistical analysis

- In practice, recognizing speech is much more complex than simply identifying phones and comparing them to stored patterns, and for a whole variety of reasons:
- Speech is extremely variable: different people speak in different ways (even though we're all saying the same words and, theoretically, they're all built from a standard set of phonemes)
- You don't always pronounce a certain word in exactly the same way; even if you did, the way you spoke a word (or even part of a word) might vary depending on the sounds or words that came before or after.
- As a speaker's vocabulary grows, the number of similar-sounding words grows too: the digits zero through nine all sound different when you speak them, but "zero" sounds like "hero," "one" sounds like "none," "two" could mean "two," "to," or "too"... and so on. So recognizing numbers is a tougher job for voice dictation on a PC, with a general 50,000-word vocabulary, than for an automated switchboard with a very specific, 10-word vocabulary containing only the ten digits.
- The more speakers a system has to recognize, the more variability it's going to encounter and the

  Bigger the likelihood of making mistakes.

Artificial neural networks

- HMMs have dominated speech recognition since the 1970s—for the simple reason that they work so well. But they're by no means the only technique we can use for recognizing speech.

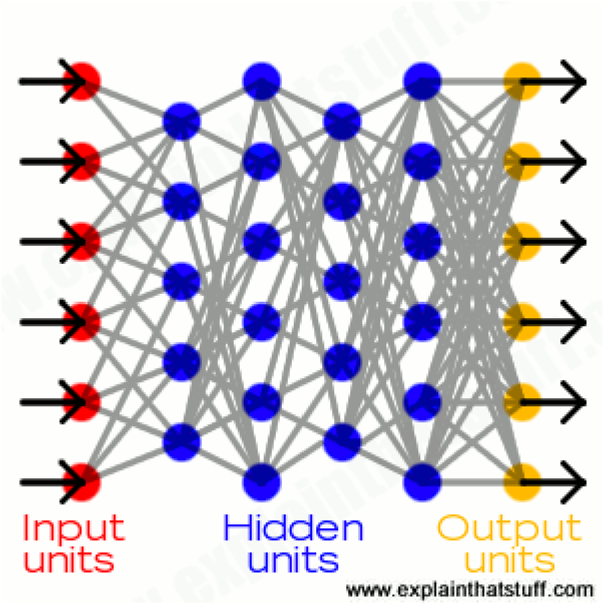
There's no reason to believe that the brain itself uses anything like a hidden Markov model. It's much more likely that we figure out what's being said using dense layers of brain cells that excite and suppress one another in intricate, interlinked ways according to the input signals they receive from our cochlea's (the parts of our inner ear that recognize different sound frequencies).

- Back in the 1980s, computer scientists developed "connectionist" computer models that could

  Mimic how the brain learns to recognize patterns, which became known as artificial neural networks (sometimes called ANNs). A few speech recognition scientists explored using neural networks, but the dominance and effectiveness of HMMs relegated alternative approaches like this to the sidelines. More recently, scientists have explored using ANNs and HMMs side by side and found they give significantly higher accuracy over HMMs used alone.

Neural networks

Neural networks are hugely simplified, computerized versions of the brain—or a tiny part of it that have inputs (where you feed in information), outputs (where results appear), and hidden units (connecting the two). If you train them with enough examples, they learn by gradually adjusting the strength of the connections between the different layers of units. Once a neural network is fully trained, if you show it an unknown example, it will attempt to recognize what it is based on the examples it's seen before.



Dynamic time warping (DTW)-based speech recognition

- Dynamic time warping is an algorithm for measuring similarity between two sequences that may vary in time or speed. For instance, similarities in walking patterns would be detected, even if in one video the person was walking slowly and if in another he or she were walking more quickly, or even if there were accelerations and deceleration during the course of one observation.
- DTW has been applied to video, audio, and graphics – indeed, any data that can be turned into a linear representation can be analyzed with DTW.

Hidden Markov models

- These are statistical models that output a sequence of symbols or quanttes. HMMs are used in

Speech recognition because a speech signal can be viewed as a piecewise stationary signal or a

Short-time stationary signal. In a short time-scale (e.g., 10 milliseconds), speech can be

Approximated as a stationary process.

- Speech can be thought of as a Markov model for many stochastic purposes.
- Another reason why HMMs are popular is because they can be trained automatically and are simple and computationally feasible to use
- In speech recognition, the hidden Markov model would output a sequence of n-dimensional real-

Valued vectors (with n being a small integer, such as 10), outputting one of these every 10

Milliseconds.

- The vectors would consist of cepstral coefficients, which are obtained by taking a Fourier transform

Of a short time window of speech and de-correlating the spectrum using a cosine transform, then

Taking the first (most significant) coefficients. The hidden Markov model will tend to have in each

State a statistical distribution that is a mixture of diagonal covariance Gaussians, which will give a Likelihood for each observed vector. Each word, or (for more general speech recognition systems),

- Each phoneme, will have a different output distribution; a hidden Markov model for a sequence of Words or phonemes is made by concatenating the individual trained hidden Markov models for the Separate words and phonemes.
- Usage in education and daily life-For language learning, speech recognition can be Useful for learning a second language. It can teach proper pronunciation, in addition to helping a person develop fluency with their speaking skills.
- Students who are blind (see Blindness and education) or have very low vision can Benefit from using the technology to convey words and then hear the computer recite Them, as well as use a computer by commanding with their voice, instead of having to Look at the screen and keyboard.
- Students who are physically disabled or suffer from Repetitive strain injury/other

Injuries to the upper extremities can be relieved from having to worry about

Handwriting, typing, or working with scribe on school assignments by using speech-to-

Text programs. They can also utilize speech recognition technology to freely enjoy searching the Internet or using a computer at home without having to physically operate a mouse and keyboard.[63]

Performance

- Error rates increase as the vocabulary size grows
- Vocabulary is hard to recognize if it contains confusable words
- Speaker dependence vs. independence:
- Isolated, Discontinuous or continuous speech
- Task and language constraints
- Read vs. Spontaneous Speech
- Adverse conditions
- Acoustical signals are structured into a hierarchy of units;

e.g. Phonemes, Words, Phrases, and Sentences

VI WHAT ARE THE POSSIBLE SOLUTIONS?

Feature analysis:

If this technique is used the user need not be trained to follow a particular accent or a particular speed in which he has to speak.

The words that have to be defined need not be confined to very few words

Simple pattern matching

- In simple pattern matching. We could have the computer talk back to the user and fulfill his requirement.
- It is similar to that of what happens with an automated machine

Statistical analysis

- The computer would have studied the behavior of the user.
- And arrived with a statistically report as to how the user would communicate with the computer.
- Limitation: not a good choice for an application with many different Users.

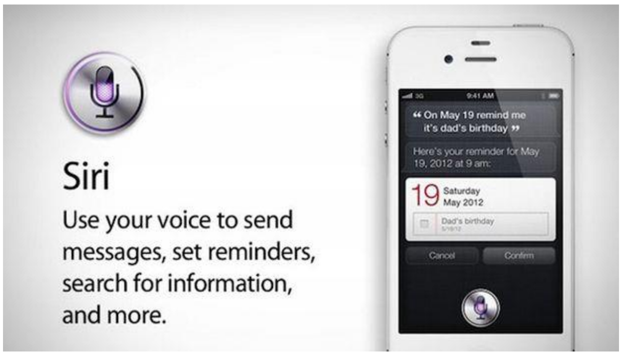
Artificial neural networks

- With this technique: The vocabulary would have been vast.
- As the users keeps searching for new unknown things. Yet there would have been a result generated by trying to connect it with an action that was performed previously.
- That is referring to the searches that were previously being performed.
- This technique can be used in order to increase the number words that could have been embedded in the application/program

Hidden Markov technique

- In this technique: only the phrase or sentence spoken in a short span of time is processed.
- So in case we have phrases that perform a particular action and the user takes a long time to finish the sentence then the appropriate result is not attained.

Mostly commonly used speech recognition software in phones


Siri
Use your voice to send messages, set reminders, search for information, and more.
On May 19 remind me it's dad's birthday
Here's your reminder for May 19, 2012 at 9 am:
19 Saturday May 2012
Dad's birthday
Cancel
Confirm

REFERENCES:

[1] Wikipedia https://en.wikipedia.org/wiki/Speech_recognition

[2] *"voice recognition, definition of". WebFinance, Inc. Retrieved February 21, 2012.* Morgan, Bourlard, Renals, Cohen, Franco (1993) "Hybrid neural network/hidden Markov model systems for continuous speech recognition. ICASSP/IJPRAI"

[3] Nuance Exec on iPhone 4S, Siri, and the Future of Speech". Tech.pinions. October 10, 2011. RetrievedNovember 23, 2011.

[4] http://ethw.org/First-Hand:The_Hidden_Markov_Model