



Recog: Artificial Intelligent Software for Speech Recognition

Dr. Rajendra Kumar Mahto

Assistant Professor

Department Of Information Technology,

Dr. Shayama Prasad Mukherjee University, Ranchi

Abstract: The software known as recog, which I created as a component of my artificial intelligence project, is the subject of this paper. The idea of speech recognition serves as its foundation. It recognizes a few words from the program and executes a specific action. The Visual Studio platform serves as its foundation. Additionally, the purpose of speech recognition, its distinction from voice recognition, its application, the artificial intelligence technique used in this project, and its selection are all explained in this paper. What different approaches might have been taken?

Keyword: artificial intelligence, speech recognition, voice recognition, ANN, HMM, Hidden Markov Model.

I. INTRODUCTION

The multidisciplinary subfield of computational linguistics known as speech recognition (SR) combines linguistics knowledge and research. computer science, electrical engineering, and other fields to create methods and technologies that allow computers and other computerized devices, like robotics and Smart Technologies, to recognize spoken language and translate it into text. Other names for it include "speech to text" (STT), "automatic speech recognition" (ASR), and "computer speech recognition."

Speaker identification, also known as voice recognition [2][3][4], refers to recognizing the speaker rather than what they are saying. In systems that have been trained on a particular person's voice, identifying the speaker can make the process of translating speech easier. It can also be used as a security measure to authenticate or verify the identity of a speaker. Speaking of technology, speech recognition has a lengthy history and multiple waves of significant inventions. Recent developments in big data and deep learning have helped the field. The proliferation of scholarly works in the field and, more significantly, the widespread industry adoption of diverse deep learning techniques in the development and implementation of speech recognition systems serve as testaments to the advancements in the field. Several of the major players in the speech industry, such as Microsoft, Google, IBM, Baidu (China), Apple, Amazon, Nuance, and IflyTek (China), have made it known that the deep learning technology at the heart of their speech recognition systems is what makes them work.

Speech Recognition: What Is It?

- The ability of a computer or program to recognize words and phrases in spoken language and translate them into a format that is readable by machines is known as speech recognition.
- You are able to utilize your voice to command your computer. You can dictate text to the computer and give commands to it that it will obey.

- In "training" (also known as "enrollment"), certain SR systems employ a single speaker who reads text or specific vocabulary into the system. The system employs voice analysis to fine-tune the recognition of that individual's speech, leading to an improvement in accuracy. Systems referred to as "speaker independent"[1] are those that do not require training. We refer to systems that rely on training as "speaker dependent".
- Applications for speech recognition include voice user interfaces for domestic appliance control, voice dialling (e.g., "Call home"), call routing (e.g., "I would like to make a collect call"), search (e.g., find a podcast where particular words were spoken), and basic data entry (e.g., entering a credit card number), setting up the
- A microphone connection to your computer is required. Once the microphone is configured, you can use voice recognition software to train your computer to recognize spoken commands and your voice by recording a voice profile. For details on how to rig your microphone. Following the setup of your microphone and voice profile, Speech Recognition can be used for the following tasks:
- Take charge of your computer. When you give spoken commands, Speech Recognition hears you out and reacts. Software can be executed and Windows can be interacted with using Speech Recognition. See Common Commands in Speech Recognition for additional details on the commands you can use with the program.
- Edit and dictate text. It is possible to use Speech organized records, such as... a radiology report), speech-to-text (Word Processors, Emails, etc.) and aircraft (often referred to as Direct Voice Input).
- Use speech recognition to transcribe text into word processing apps or complete online forms in a web browser. Additionally, speech recognition can be used to edit text on your computer. See Dictate text using Speech Recognition for additional details on text dictation.

Speech recognition and voice recognition are not the same.

- When voice recognition is used to identify the physical person speaking, it usually ignores the language and meaning. To detect words, speech recognition eliminates individual differences.
- Speech recognition operates without regard to language.
- The recognition of speech varies with language.
- What makes speech recognition software necessary?
- quantity of programs available to users with and without disabilities.
- Speech-to-text has been utilized to assist struggling authors in increasing their output.
- Additionally, to provide people with physical disabilities an alternative method of computer access. Additional uses for speech recognition include learning foreign languages. devices for the blind that are voice activated, as well as numerous well-known mainstream technologies.
- automated phone directories and menus, voice-activated phone calls on our mobile devices, and Smartphones with built-in voice commands are just a few instances.
- Voice recognition is a daily tool used by legal and medical professionals to dictate notes and write down any pertinent information. The technology is now being used in military applications, navigation systems, and automobile Voice command devices in "smart" homes, speech recognition (Ford SYNC), and Playing video games like End War lets the user command their troops with just their voice.
- The benefits and drawbacks of speech recognition technology lessens the complexity of time lowers the user's overhead. Interface that is user-friendly and interactive.
- lessens the complexity of space. used to look up a file or application remotely. Users with limited knowledge of grammar can still access the files and programs.
- The user must adhere to the software's voice instructions in order to prevent incorrect output or errors.

II. RECOG:

Applications currently in use

- Amazon
- Gmail
- Facebook
- Flipkart
- VLC
- Run it open
- Chrome
- Excel
- Word Power point
- Twitter
- Recognize

Form-integrated applications

- Recognize:
- Webcam
- Player for media
- picture viewer
- Player of movies
- Maps
- Interpreter
- Google lookup
- Lookup on YouTube
- The weather

Features

- It launches any of the specified programs without requiring the use of a keyboard or mouse.
- The rodent is merely utilized to either disable or enable speech recognition.
- The speech recognition engine and synthesizer identify everything the user says.
- The relevant block, which opens the necessary software or performs an action, is then executed after the text has been compared to each of the switch case conditions.
- The current form is closed with the close command.

The project's objective

- Applications can be accessed and opened without a mouse or keyboard. Take photos without pressing the shutter release button.
- Watch a movie without navigating or causing a mouse click to occur.
- Simply speak to the computer to play a song, view a picture, or watch a movie without having to look for it, or click the emblem. Smartphone "personal assistant apps" include Microsoft's Cortana, Apple's Siri, and Google's now.

Platform that the project is constructed upon

- Microsoft Visual studio

Other systems that can be used to develop it

- VB
- PHP
- MATLAB

Restrictions

- This application requires the Visual Studio software to operate.
- It is not a stand-alone piece of software. The speech recognition process is initially initiated by using the mouse to both enable and disable the speech recognition engine and synthesizer. Not every application has an embedded version. Given that Google APIs were necessary.

Applications

- automobile systems
- Medical records related to health care
- Military high-performance fighter aircraft: commanding automated target handover systems, establishing navigational systems, and operating communication radios.
- Telephony and other domains: SR is becoming widely used in the telephony industry. The realm of computer simulation and gaming is expanding in popularity. Even so, word processing is highly integrated into general personal computing. However, the anticipated [by whom?] increases in use of ASR in the document production field have not materialized.

III. TEMPLATE MATCHING IS THE TECHNIQUE USED TO SOLVED THE PROBLEM

- When applied correctly, template matching is the most accurate and simplest method, but it also has the most drawbacks.
- The user speaks a word or phrase into a microphone as the first step.
- An "analog-to-digital (A/D) Converter" digitizes the electrical signal coming from the microphone and stores it in memory.
- The computer compares the voice input with a digital voice sample in an attempt to ascertain the "meaning" of this voice input. or model with a predetermined meaning.
- This method bears a strong resemblance to the conventional keyboard-based command inputs.
- The input template is contained in the program, which tries to match it with the Real input with a basic conditional expression.
- "Speaker dependent." systems of this kind, and recognition accuracy can be roughly 98%.

IV. WHY USE THIS METHOD?

- The program was simple to put into place.
- Each word has a specific function to fulfil.
- Only a limited number of words are selected, and those are assigned a specific task to complete.
- Using this method makes the search easier because there are numerous forms available.
- The terms used to carry out an operation vary depending on the form, and the quantity of Each form has different operations.
- The purpose of this approach is to shorten the time that the current system takes to open a file, retrieve a weather report, etc.
- The primary goal of the suggested system is to shorten the time needed to complete specific tasks.
- The suggested software uses less memory than the current software and can be used to open videos, audio files, weather reports, and websites like Google, YouTube, Twitter, and others simply by speaking commands.

V. WHAT ARE THE OPTIMAL METHODS?

ANALYSIS OF FEATURES:

Speech recognition that is "speaker-independent" is typically the result of this technique.

- This method uses "Fourier transforms" to process the voice input first, rather than attempting to find an exact or nearly exact match between the actual voice input and a voice template that has already been stored. or "linear predictive coding (LPC)", subsequently looks for distinguishing features that the expected inputs and the actual digitized voice input have in common.
- Numerous speakers will share these characteristics, so every new user won't need to train the system.
- Accents, as well as variations in delivery speed, pitch, loudness, and inflection, are among the speech differences that the speaker-independent method can handle but that pattern matching would not be able to.
- The variety of accents and inflections used by speakers of different nationalities has proven to be one of the biggest obstacles in the extremely challenging field of speaker-independent speech recognition.

Basic pattern recognition

- If you've ever called an automated call center and gotten a computerized response, you've probably encountered it. switchboard.

- These kinds of systems are frequently available from utility companies, and banks occasionally use them to automate routine tasks like checking accounts, placing orders for statements, leaving meter readings, and so forth.
- It's as easy as dialling a number, waiting for a recorded voice to answer, and then entering or speaking your account number and pressing additional keys to choose what you want to do.
- Importantly, you are never required to do anything more complicated than parse a sentence—that is, break up a spoken string of sounds into individual words and determine their structure—or even attempt to comprehend the computer on the other end. All you ever have to do is select one option from a very small list.

Analytical statistics

- For a multitude of reasons, speech recognition in practice is far more complicated than just identifying phones and matching them to stored patterns.
- Speech is incredibly variable; even though we all say the same words and, in theory, they are all constructed from a standard set of phonemes, people speak in a variety of ways.
- You don't always pronounce words exactly the same way; even if you did, your speech pattern wouldn't depend on the sounds or words that came before or after, a word (or even a portion of a word) may change.
- The number of words that sound similar increases with a speaker's vocabulary: the numbers zero through nine all sound different when spoken, but "zero" sounds like "hero," "one," and so on. "none," "two" could imply "two," "to," "too," and so forth. Therefore, voice dictation on a PC with a broad 50,000-word vocabulary finds it harder to recognize numbers than an automated switchboard with a very narrow 10-word vocabulary that only includes the ten digits.
- A system will face greater variability the more speakers it must identify, and the increased probability of errors.

Artificial Neural Networks

- Since the 1970s, HMMs have dominated speech recognition for the straightforward reason that they are so effective. However, these are by no means the only methods available to us for speech recognition.
- It is not plausible to assume that the brain functions in any way like a hidden Markov model. It's far more probable that we have delicate layers of brain cells that respond to input signals from our cochleas—the parts of our inner ear that detect different sound frequencies—excite and suppress one another in complex, interconnected ways to interpret what is being said.
- Computer scientists created "connectionist" computer models in the 1980s that could Artificial neural networks, also known as ANNs, are designed to mimic how the brain learns to recognize patterns. A few researchers in speech recognition experimented with neural networks, but alternative methods such as this one were pushed to the periphery due to the popularity and efficiency of HMMs. Scientists have recently experimented with combining ANNs and HMMs side by side and discovered that ANNs provide noticeably better accuracy than HMMs alone.

Networks of neurons

Neural networks are incredibly the brain or a small portion of it with inputs (where information is fed in), outputs (where results appear), and hidden units (connecting the two) is represented by simplified, computerized versions. They learn by gradually changing the strength of the connections between the various layers of units if you provide them with enough examples to train them. When a neural network is fully trained, it will try to identify an unknown example by using the examples it has previously seen as a guide.

Speech recognition using dynamic time warping (DTW) technique

- An algorithm for comparing the similarity of two sequences that might differ in speed or duration is called dynamic time warping. For example, similarity in gait would be identified even if the subject walked slowly in one video. and if, during one observation, there were accelerations and decelerations, or even if they were walking more swiftly in another.
- Video, audio, and graphics have all been analyzed using DTW; in fact, any type of data that can be transformed into a linear representation can be used for analysis.

Markov models that are hidden

- These statistical models produce a series of quanttes, or symbols. HMMs are employed in Considering that a speech signal can be seen as either a piecewise stationary signal or a Short-time stationary signal. Within a brief duration (such as 10 milliseconds), voice can be approximated as a process that is stationary.
- For many stochastic purposes, speech can be considered a Markov model.
- HMMs' popularity is also due to their automatic training capabilities, ease of use, and computational viability.
- The hidden Markov model in speech recognition would produce an n-dimensional real-Valued vectors, producing one of these every ten times (where n is a small integer, like 10).

Achievement

- As vocabulary increases, error rates rise.
- It is difficult to identify vocabulary if it contains unclear terms.
- Speaker independence versus reliance:
- Speech that is isolated, interrupted, or continuous Task and linguistic limitations
- Read Aloud vs. Unplanned Speech adverse circumstances A hierarchy of units is used to structure acoustic signals;
- Words, phrases, sentences, and phonemes, for example

VI. WHAT ARE THE POSSIBLE SOLUTION?

Analyzing features:

- The user does not need to be trained to speak with a specific accent or at a set pace if this technique is applied.
- There is no need to limit the words that must be defined to a small number of words.
- basic pattern recognition in straightforward pattern matching. We could satisfy the user's request by having the computer respond to him.
- It is comparable to what occurs when a machine is automated.

Analytical statistics

The user's behaviour would have been examined by the computer and showed up with a statistical report detailing the user's expected interface with the computer.

Limitation: not ideal for a situation where a lot of different People.

Artificial Neural Networks

Using this method: There would have been a vast vocabulary.

As the user continues to look for new, undiscovered items. However, attempting to link it to a previously completed action would have produced a result.

That's a reference to the earlier searches that were carried out.

Markov technique with hidden variables

This method processes only the phrase or sentence that is spoken in a brief amount of time.

Therefore, the intended outcome is not achieved if we have phrases that carry out specific actions and the user takes a long time to finish the sentence.

Generally speaking, phones with speech recognition software

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. Retrieved November 23, 2011.
- [4] http://ethw.org/First-Hand:The_Hidden_Markov_Model

