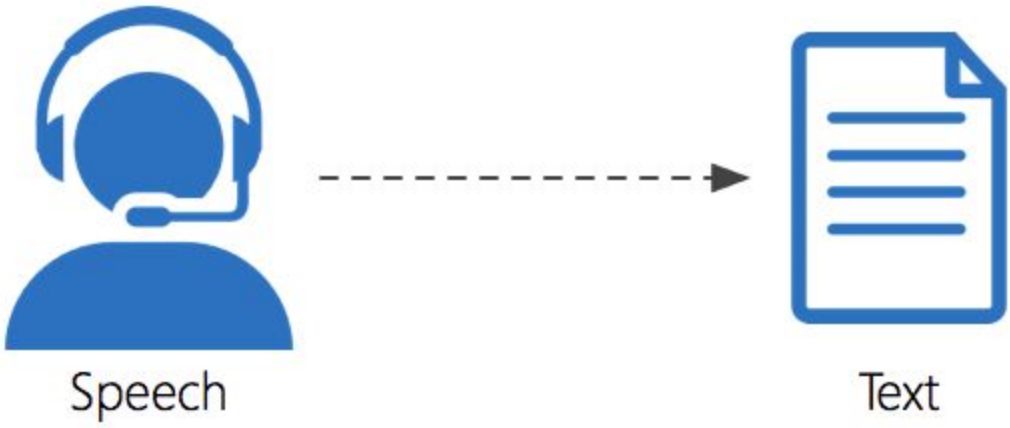
SPEECH TO TEXT



Arunkumar Panneerselvam

13.10.2020

INTRODUCTION

Speech is the most common means of communication and the majority of the population in the world relies on speech to communicate with one another. Speech recognition systems basically translate spoken languages into text. There are various real-life examples of speech recognition systems. For example, Apple SIRI, Google Voice Assistant which recognizes the speech and truncates it into text.

HISTORY [1]

1950S AND 60S

The first speech recognition systems were focused on numbers, not words. In 1952, Bell Laboratories designed the "Audrey" system which could recognize a single voice speaking digits aloud. Ten years later, IBM introduced "Shoebox" which understood and responded to 16 words in English.

Across the globe other nations developed hardware that could recognize sound and speech. And by the end of the '60s, the technology could support words with four vowels and nine consonants.

1970S

Speech recognition made several meaningful advancements in this decade. This was mostly due to the US Department of Defense and DARPA. The Speech Understanding Research (SUR) program they ran was one of the largest of its kind in the history of speech recognition. Carnegie Mellon's "Harpy' speech system came from this program and was capable of understanding over 1,000 words which is about the same as a three-year-old's vocabulary. Also significant in the '70s was Bell Laboratories' introduction of a system that could interpret multiple voices

1980S

The ‘80s saw speech recognition vocabulary go from a few hundred words to several thousand words. One of the breakthroughs came from a statistical method known as the “Hidden Markov Model (HMM)”. Instead of just using words and looking for sound patterns, the HMM estimated the probability of the unknown sounds actually being words.

1990S

Speech recognition was propelled forward in the 90s in large part because of the personal computer. Faster processors made it possible for software like Dragon Dictate to become more widely used.

BellSouth introduced the voice portal (VAL) which was a dial-in interactive voice recognition system. This system gave birth to the myriad of phone tree systems that are still in existence today.

2000S

By the year 2001, speech recognition technology had achieved close to 80% accuracy. For most of the decade there weren’t a lot of advancements until Google arrived with the launch of Google Voice Search. Because it was an app, this put speech recognition into the hands of millions of people. It was also significant because the processing power could be offloaded to its data centers. Not only that, Google was collecting data from billions of searches which could help it predict what a person is actually saying. At the time Google’s English Voice Search System included 230 billion words from user searches.

2010S

In 2011 Apple launched Siri which was similar to Google's Voice Search. The early part of this decade saw an explosion of other voice recognition apps. And with Amazon's Alexa, Google Home we've seen consumers becoming more and more comfortable talking to machines.

Today, some of the largest tech companies are competing to herald the speech accuracy title. In 2016, IBM achieved a word error rate of 6.9 percent. In 2017 Microsoft usurped IBM with a 5.9 percent claim. Shortly after that IBM improved their rate to 5.5 percent. However, it is Google that is claiming the lowest rate at 4.9 percent.

THE FUTURE

The technology to support voice applications is now both relatively inexpensive and powerful. With the advancements in artificial intelligence and the increasing amounts of speech data that can be easily mined, it is very possible that voice becomes the next dominant interface.

At Sonix, we can thank the many companies before us that have propelled speech recognition to where it is today. We automate transcription workflow and make it fast, easy, and affordable. We couldn't do that without the amazing work that has been done before us.

REQUIREMENTS

1. Python 3.7.3
2. “PyAudio”Library

   This Python library is used for audio input/output operations through the microphone and speaker. It will help to get our voice through the microphone.

   Several Speech recognition engine/API support[2]:

   - CMU Sphinx (works offline)
   - Google Speech Recognition
   - Google Cloud Speech API
   - Wit.ai
   - Microsoft Bing Voice Recognition
   - Houndify API
   - IBM Speech to Text
   - Snowboy Hotword Detection (works offline)

3. I’m going to use Google Speech-To-Text API. It isn't free, however. It is free for speech recognition for audio less than 60 minutes.

PROCEDURE

Step 1: Execute the below command in your command prompt to install a "Speech Recognition" API in Python. Before installation you will verify your Python version, which is "Python 3.7.3"

pip install SpeechRecognition

Step 2: We need to install PyAudio library which used to receive audio input and output through the microphone and speaker. Basically, it helps to get our voice through the microphone. (refer in such case you got error)

pip install PyAudio

Step 3 : Run the code (https://github.com/akpgaa/Speech-to-Text-in-Python.git)

Note : Audio file supports by speech recognition: wav, AIFF, AIFF-C, FLAC.

REFERENCES

1. https://sonix.ai/history-of-speech-recognition
2. https://pypi.org/project/SpeechRecognition/
3. https://www.kdnuggets.com/2020/06/easy-speech-text-python.html
4. https://morioh.com/p/e0e5eb85a514