Apple Voice Recognition

Hey Siri,
How do you identify my voice?

T,

Learn

#1

# Introduction

___

Apple Computers, Inc. was founded on April 1, 1976, by college dropouts Steve Jobs and Steve Wozniak, who brought to the new company a vision of changing the way people viewed computers. Jobs and Wozniak wanted to make computers small enough for people to have them in their homes or offices. Simply put, they wanted a computer that was user-friendly.

Keeping that vision in mind Apple pioneered their way through the computer industry—not once, but multiple times throughout its existence. One great innovation was, Siri. The name Siri means "beautiful woman who leads you to victory". It is a digital personal assistant, integrated within Apple device operating systems, that enables Apple device users to get answers to questions, check the weather, confirm flights, perform searches, answer questions, complete actions, send a message and much more.

It is the first virtual personal assistant, arose from decades of SRI research in artificial intelligence (AI). In April 2010, Apple acquired Siri, and in October 2011, Siri was unveiled as an integrated feature of the Apple iPhone 4S.
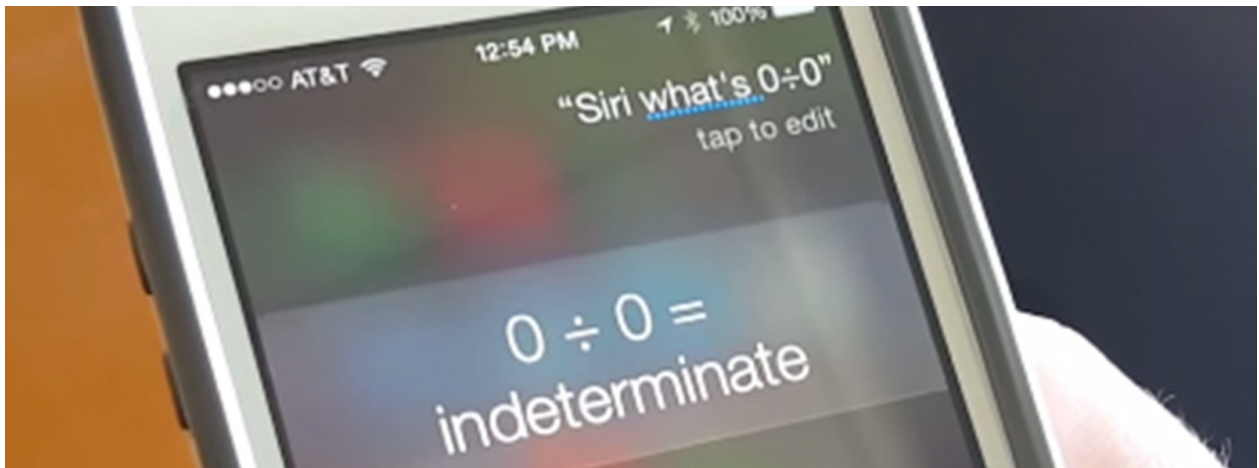
T, Learn
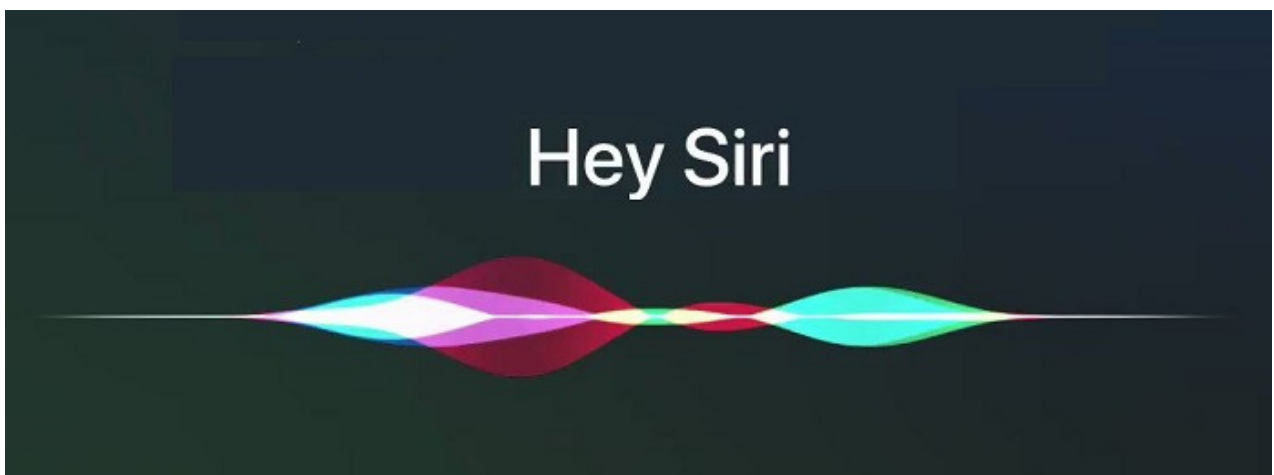
How does Siri works?

Upon receiving your request, Siri records the frequencies and sound waves from your voice and translates them into a code. Siri then breaks down the code to identify particular patterns, phrases, and keywords. This data gets input into an algorithm that sifts through thousands of combinations of sentences to determine what the inputted phrase means. This algorithm is complex enough that it is capable of working around idioms, homophones and other literary expressions to determine the context of a sentence.

Once Siri determines its request, it begins to assess what tasks needs to be carried out, determining whether or not the information needed can be accessed from within the phone's data banks or from online servers. Siri is then able to craft complete and cohesive sentences relevant to the type of question or command requested.

Overall, Siri is based on large-scale Machine Learning systems that employ many aspects of data science.

Amazing isn't it? In our technology-heavy world, we tend to take things around us for granted. However, when we try to uncover them, we realize that there is a lot of technological magic happening behind the scene.

T,
Learn

Voice Vs Speech Recognition

Parameters	Speech Recognition	Voice Recognition
Definition	Speech recognition is the ability of any machine or electronic device to identify and understand the spoken words. The words get converted into a machine-readable format.	Voice recognition is an innovative technology that allows people to interact with electronic devices by interacting with them. It aims at identifying the speaker.
Number of Applications	Speech recognition is also simplifying the tasks of many business domains. It has proved quite handy in hands-free computing and automatic translations.	Voice recognition caters to a wide range of industries and has become useful in eliminating the user verification process.
How Does it Work	The speech recognition first analyzes the sound and then filters it clearly to understand what the person is speaking. It then digitizes the sound into a readable format to analyze its meaning.	The voice recognition takes the voice sample of the person and then converts it into a digital format. It creates a voice template or print.
Evaluation of Program	Speech recognition is evaluated on two primary factors - speed and accuracy.	Voice recognition is also evaluated on two factors - speaker verification and speaker identification.
Key Benefits	Speech recognition is a time and money saving technology, which provides both speed and accuracy. It is easy to install on various devices.	Voice recognition also provides an array of benefits such as improved flexibility, enhanced productivity, reusable data, and much more.
Focus	Speech recognition usually focuses on vocabulary of the speaker. The words are then translated into digital format.	The main focus of voice recognition is to identify the biometric aspects of the speaker. It analyzes the tempo, pitch, and timbre of the voice.
Examples	Some examples of speech recognition include Google Now, Google Cloud Speech API, Google Docs Voice Typing, and more.	Examples of voice recognition include automated phone systems, digital assistants, Google voice, and car Bluetooth, etc.

T,
Learn

Technology behind Voice Identification

Voice identification technology captures and measures the physical qualities of a person's voice when speaking as well as the unique biological parameters that combine to produce that voice.
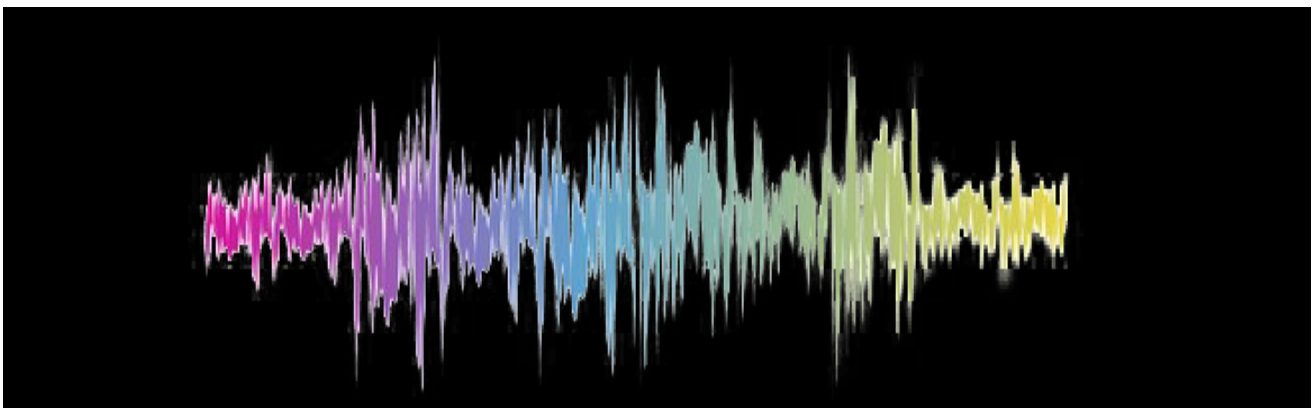
These parameters Include:

Pitch - Pitch is an important perceptual dimension by which listeners discriminate and categorize voice quality. It affects the perceived brightness of the sound, and brightness may be one of several perceptual features of a sound used by listeners to distinguish one voice quality from another.

Intensity - The increased vocal intensity results from a greater resistance by the vocal folds to increased airflow. The vocal folds are blown wider apart, releasing a larger puff of air that sets up a sound pressure wave of greater amplitude.

Dynamics - Within-person variability in our vocal signals is substantial: we volitionally modulate our voices to express our thoughts and intentions or adjust our vocal outputs to suit a particular audience, speaking environment, or situation.
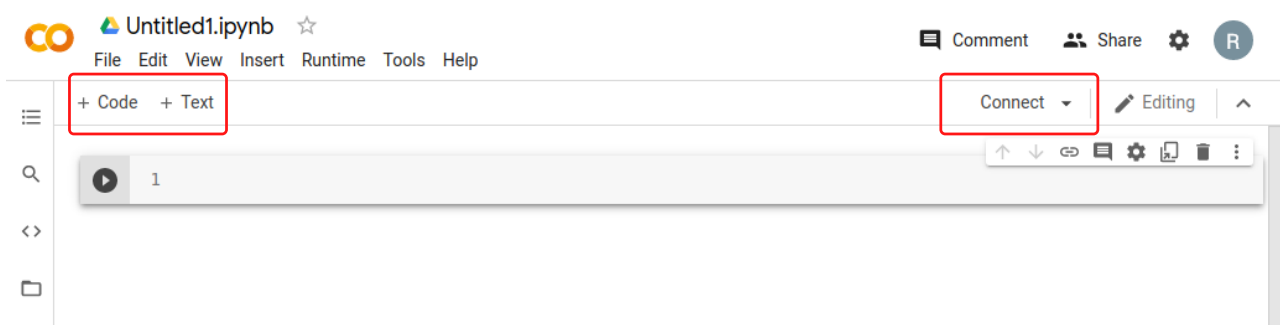
Introduction to Colab

Nearly all machine learning and deep learning algorithms require good hardware. What if ... you don't have good hardware? Should you drop your dream to be a data scientist? No, there's an alternative Let us introduce you to Colab.

Colab is a service provided by Google which lets you access a virtual machine hosted on google servers. These virtual machines have dual-core Xeon processors, with 12GB of RAM. You can even use GPU for your neural networks. Colab is an interactive python notebook (ipynb), which means that with writing python code, you can also write normal text, include images.
To create a new colab notebook, just go to "https://colab.research.google.com", and create a new notebook. You will get something like below

You can connect to runtime by clicking the "Connect" button in the right corner. You can add a new Code cell, or text cell using respective buttons in the toolbar.
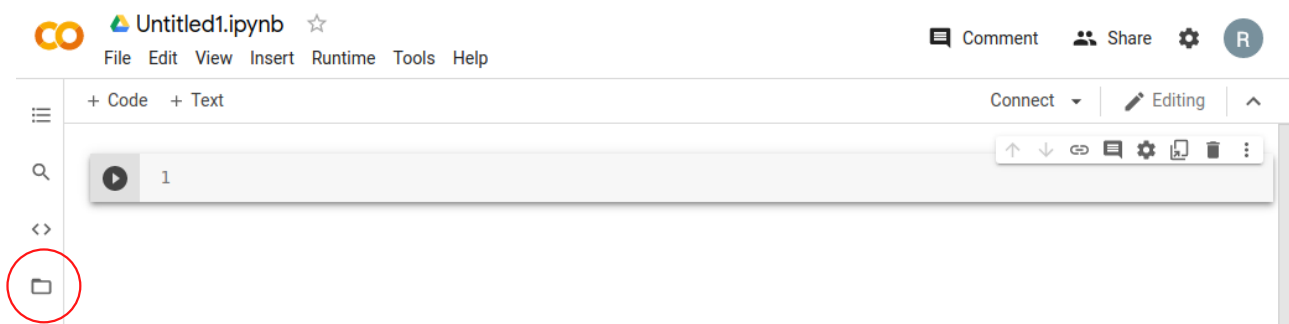
T,
Learn

Introduction to Colab

Uploading files

Sometimes you need to use a file from your PC. For that, you can upload the required files to google colab. Colab provides 100 GB in a colab session. If you want to access some files from your drive, you can even do so by connecting the drive to colab.

To upload something, open file pane from left toolbar.

Now, just upload the file using first button, and you can also mount google drive using last button

T,
Learn
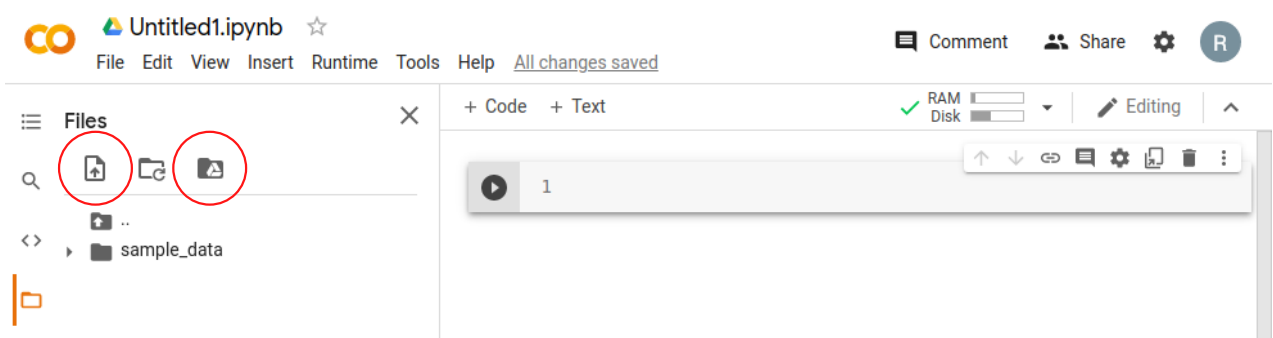
#3.1

Code

Install & Import

Our code majorly make use of different libraries and tool kits. Let us understand what each library do.

speaker-verification-toolkit

This module contains some tools to make a simple speaker verification.

```
!pip install speaker-verification-toolkit

import speaker_verification_toolkit.tools as svt
```

Numba

Within-person variability in our vocal signals is substantial: we volitionally modulate our voices to express our thoughts and intentions or adjust our vocal outputs to suit a particular audience, speaking environment, or situation.

```
!pip install numba==0.48
```

T,
Learn

Code

Import Librosa

Librosa is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems.

```
import librosa
```

Upload audio files

Take any two audios of the same person and one audio of some one else. Make sure the audio files have .wav extension. Librosa only supports .wav extension.

```
data1, sr = librosa.load('file path', sr=16000, mono=True)
data2, sr = librosa.load('file path', sr=16000, mono=True)
```

Remove Noise

Cut off silence parts from the signal audio data. Doesn't work with signals data affected by environment noise.

```
data1 = svt.rms_silence_filter(data1)
data2 = svt.rms_silence_filter(data2)
```

T,
Learn

Code

Extract unique features

The Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method is a leading approach for speech feature extraction and current research aims to identify performance enhancements.

```
data1 = svt.extract_mfcc(data1)
data2 = svt.extract_mfcc(data2)
```

Calculate difference b/w audios

Compute the distance between sample1 and sample2.

```
print(
    'The difference between voice1 and voice2 is',
    svt.compute_distance(data1,data2)
)
```

Upload and repeat steps for Sample 3

```
data3, sr = librosa.load('file path', sr=16000, mono=True)
data3 = svt.rms_silence_filter(data3)
data3 = svt.extract_mfcc(data3)
svt.compute_distance(data3, data2)
```

T,
Learn

Code

Finding Nearest Voice

Find the nearest voice data based on this voice sample. Could be used to make the naive Accept/Reject decision.

```
svt.find_nearest_voice_data([ data3, data1], data2)
```

The output will be the sample who is the nearest.

T,
Learn

T,
Learn

Knowledge **Discovery** through **Brand** Stories