

Reconocimiento de Voz

(agosto de 2022)

Angel Moreno, Ivys Chávez, y Yadel García

Resumen

Un sistema de reconocimiento de voz es la capacidad que presenta un ordenador para recibir los datos de voz de un usuario, transformar la señal en código binario, el cual es asimilado por la computadora y luego establécela comunicación hombre-máquina necesaria para resolver diferentes problemas que requieran para su resolución la utilización de este método. Desde el punto de vista social se emplea como una herramienta útil y necesaria para personas con discapacidad física (carencia en sus extremidades superiores); además de agilizar la búsqueda de información propia o a través de la red para cualquier usuario que posea un ordenador con esta herramienta tecnológica. Esta tecnología podría estar convirtiéndose en un estándar en los nuevos dispositivos, pero su precisión es finalmente lo que determina si realmente se convierte en un recurso sin posibilidad de uso. En este proyecto se utiliza el reconocimiento de voz para escribir un texto.

I. INTRODUCCION

El reconocimiento de voz, como su nombre indica, se refiere al reconocimiento automático del habla humana. El reconocimiento de voz es una de las tareas más importantes en el dominio de la interacción entre humanos y computadoras. Si alguna vez ha interactuado con Alexa o le ha pedido

a Siri que complete una tarea, ya ha experimentado el poder del reconocimiento de voz.

El reconocimiento de voz tiene varias aplicaciones que van desde la transcripción automática de datos de voz (como mensajes de voz) hasta la interacción con robots a través del habla.

Desarrollo

El reconocimiento de voz es **la capacidad de una máquina o programa para identificar palabras y frases en lenguaje hablado y convertirlas a un formato legible por máquina.**

El software de reconocimiento de voz rudimentario tiene un vocabulario limitado de palabras y frases, y solo puede identificarlas si se hablan con mucha claridad. Un software más sofisticado tiene la capacidad de aceptar el habla natural.

El primer intento registrado en la tecnología de reconocimiento de voz se remonta a 1,000 AD a través del desarrollo de un instrumento que supuestamente podría responder «sí» o «no» a preguntas directas.

Aunque este experimento técnicamente no involucró el procesamiento de voz de ninguna forma, la idea detrás de esto sigue siendo parte de la base de la tecnología de reconocimiento de voz: usar el lenguaje natural como entrada para desencadenar una acción.

Siglos después, los laboratorios Bell trabajaron para desarrollar «Audrey», un sistema capaz de reconocer los números del 1 al 9 pronunciados por una sola voz.

Más tarde, IBM desarrolló un dispositivo que podía reconocer y diferenciar entre 16 palabras habladas.

Estos éxitos provocaron una mayor prevalencia de empresas de tecnología centradas en tecnologías relacionadas con el habla. De hecho, incluso el Departamento de Defensa quería participar en la acción. Lento pero seguro, los desarrolladores avanzaron hacia el objetivo de permitir que las máquinas entiendan y respondan cada vez más a nuestros comandos verbalizados.

La historia de la tecnología de reconocimiento de voz ha sido larga y sinuosa. Sin embargo, los asistentes de voz actuales como Amazon Alexa, Google Voice, Microsoft Cortana y Siri de Apple no se encontrarían dónde están sin los primeros pioneros que simplificaron el camino.

Gracias a la integración de nuevas tecnologías, como el procesamiento basado en la nube, así como los proyectos de recopilación de datos en curso, estos sistemas de voz han mejorado continuamente su capacidad de «escuchar» y comprender una mayor variedad de palabras, idiomas y acentos.

¿Como funciona?

Rodeado de teléfonos inteligentes, automóviles inteligentes, electrodomésticos inteligentes, asistentes de voz y más, es fácil dar por sentado **cómo funciona realmente la tecnología de reconocimiento de voz**.

¿Por qué?

Porque la simplicidad de poder hablar con asistentes digitales es engañosa. El reconocimiento de voz es realmente increíblemente complicado, incluso ahora.

Piensa en cómo un niño aprende un idioma.

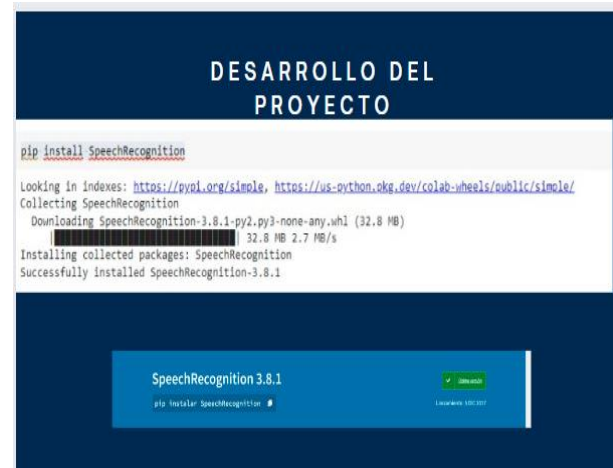
Desde el primer día, escuchan palabras que se usan a su alrededor. Los padres hablan con sus hijos y, aunque el niño no responde, absorben todo tipo de señales verbales; entonación, inflexión y pronunciación; su cerebro forma patrones y conexiones basadas en cómo sus padres usan el lenguaje.

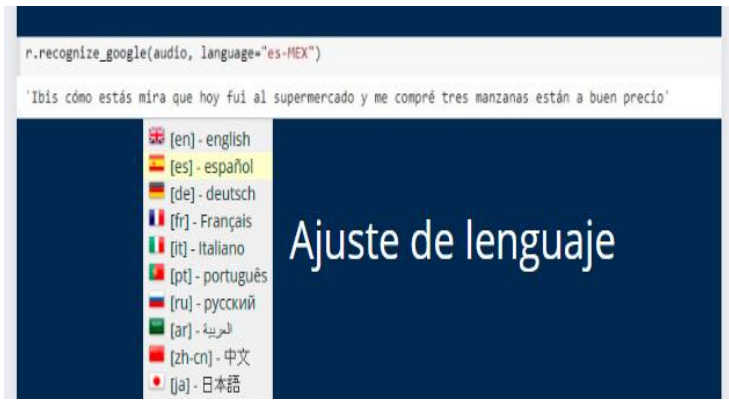
Aunque parezca que los humanos están programados para escuchar y comprender, en realidad hemos estado entrenando toda nuestra vida para desarrollar esta llamada habilidad natural.

La tecnología de reconocimiento de voz funciona esencialmente de la misma manera. Mientras que los humanos han refinado nuestro proceso, todavía estamos descubriendo las mejores prácticas para las computadoras. Tenemos que entrenarlos de la misma manera que nuestros padres y maestros nos capacitaron. Y esa capacitación involucra mucho pensamiento innovador, mano de obra e investigación.

A partir de mayo de 2017, los algoritmos de **inteligencia artificial** de Google han alcanzado una tasa de precisión de palabras del 95% para el idioma inglés.

Desarrollo del Proyecto





Resultado

```
salu2 = sr.AudioFile("audio5.wav")
with salu2 as source:
    audio = r.record(source)
```

```
r.recognize_google(audio)
```

```
'you look at 19 units of molecular'
```

```
with salu2 as source:
    r.adjust_for_ambient_noise(source)
    audio = r.record(source)
```

```
r.recognize_google(audio)
```

```
'salaam alaikum'
```

```
r.recognize_google(audio, language="es-PA")
```

```
'es lo que mantiene unidas a las moléculas esta pregunta es clave en la formación el desarrollo y entendimiento de los compuestos químicos es por ello que este tema se centra en el estudio de las fuerzas intermoleculares dentro de una molécula los átomos están unidos mediante fuerzas intramoleculares enlace iónico metal es o covalentes principalmente estas son las fuerzas que se deben vencer para que se produzca un cambio químico son estas fuerzas por tanto las que determinan las propiedades químicas de las sustancias'
```

IX. CONCLUSIÓN

Logramos aprender que realizar un proyecto usando la conversión de archivos de audio en texto utilizando la API SpeechRecognition en Python y usando google colab Podemos convertir un audio en un texto, en el que Podemos utilizar diferentes idiomas para crear un texto a partir de un audio.