**Antecedentes de Reconocimiento de Voz**

La tecnología de reconocimiento de voz es algo con lo que se ha soñado y trabajado durante décadas.

Desde el pitido de R2-D2 en Star Wars hasta la voz incorpórea pero conmovedora de Samantha en Ella, los escritores de ciencia ficción han tenido un papel muy importante en la construcción de expectativas y predicciones sobre cómo podría ser el reconocimiento de voz en nuestro mundo.

**¿Que es el reconocimiento de voz y cómo funciona?**

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.

**¿Cómo 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.

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.

**El reconocimiento de voz según el canal de comunicación**

El reconocimiento de voz robusto trata con desajustes entre entrenamiento y pruebas. Frente a una amplia gama de condiciones adversas inesperadas, los sistemas deben ser mejorados para hacer frente a las variaciones que provienen del altavoz, el tipo, dirección y posición del micrófono, el canal de transmisión y el entorno acústico.

La variabilidad relacionada con el hablante es uno de los principales factores que influyen en los sistemas de reconocimiento de voz actuales.

Los grupos de edad son algunos ejemplos de variaciones acústicas relacionadas con el hablante. Entre ellos, fenómenos inducidos por el estrés, diferencias de grupos de edad y producciones de habla no nativas, constituyen desafíos reales para el reconocimiento de voz que, desafortunadamente, funcionan razonablemente bien solo para condiciones cuidadosamente seleccionadas.

Los micrófonos actúan como filtros lineales en la señal de voz y deben tener en cuenta los diferentes grados de pendiente espectral

dependiendo de las características del micrófono. Este tipo de distorsión está involucrado con la señal del habla. Como la distancia del micrófono al que habla es a menudo diferente, también se observa variación.

**El reconocimiento de voz de Google (programas y apps)**

A fines de 2017, Google contaba con una tasa de precisión de palabras del 95% para el inglés de los Estados Unidos; el más alto de todos los asistentes de voz actualmente disponibles. Esto se traduce en una tasa de error de palabras del 4.9%, lo que hace que Google sea el primero del grupo en caer por debajo del umbral del 5%.

En lo que algunos llaman un intento de contraatacar en Amazon, Google ha lanzado muchos productos inquietantemente similares a Amazon. Por ejemplo, Google Home recuerda a Echo de Amazon, y Google Home Mini de Amazon Echo Dot.

Más recientemente, Google también anunció algunas asociaciones nuevas y clave con compañías como Lenovo, LG y Sony para lanzar una línea de «pantallas inteligentes» con el Asistente de Google, que una vez más parece «hacerse eco» de la imagen de Echo Show de Amazon.

**SeechRecognition de Google**

Biblioteca para realizar reconocimiento de voz, con soporte para varios motores y APIs, en línea y fuera de línea.

Soporte de API/motor de reconocimiento de voz:

* CMU Sphinx (funciona sin conexión)
* Reconocimiento de voz de Google
* API de Google Cloud Speech
* Wit.ai
* Reconocimiento de voz de Microsoft Bing
* API de Houndify
* Voz a texto de IBM
* Snowboy Hotword Detección (funciona sin conexión)