Computación Blanda – Agosto de 2020. Universidad Tecnológica de Pereira – Facultad de Ingenierías. Sistemas y Computación 1

Reconocimiento de Voz

Voice Reconigtion

Autor: Carlos Eduardo Hincapié López

*IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia*

Correo-e: carlos.hincapie@utp.edu.co

*Resumen*— En el siguiente documento se desea plantear el funcionamiento de un programa que reconozca diferentes voces, aprovecha el micrófono para poder grabar una muestra de audio que se esté reproduciendo. Una huella digital acústica se crea a partir de la muestra y se compara con una lista de audios en una base de datos predefinida para encontrar coincidencias. Una vez hecha la relación, el usuario puede recibir información tal como el nombre de la persona a la cual le pertenece la voz.

*Palabras clave— Shazam, Fourier, música, Transformada, Huella*

1. Introducción

Un sistema de reconocimiento de voz es una herramienta computacional capaz de procesar la señal de voz emitida por el ser humano y reconocer la información contenida en ésta, convirtiéndola en texto o emitiendo órdenes que actúan sobre un proceso.

Los sistemas comerciales han estado disponibles desde 1990. A pesar del aparente éxito de estas tecnologías, muy pocas personas utilizan el sistema del reconocimiento del habla en sus computadoras. Parece ser que muchos de los usuarios utilizan el ratón y el teclado para guardar o redactar documentos, porque les resulta más cómodo y rápido a pesar del hecho de que todos podemos hablar a más velocidad de la que tecleamos. Sin embargo, mediante el uso de ambos, el teclado y el reconocimiento del habla, nuestro trabajo sería mucho más efectivo. Este sistema donde está siendo más utilizado es en aplicaciones telefónicas: agencias de viajes, atención al cliente, información etc. La mejora de estos sistemas de reconocimiento del habla ha ido aumentando y su eficacia cada vez es mayor.

1. Usos y aplicaciones

Aunque en teoría cualquier tarea en la que se interactúe con un ordenador puede utilizar el reconocimiento de voz, actualmente las siguientes aplicaciones son las más comunes:

* Dictado automático: El dictado automático es, hasta hoy, el uso más común de las tecnologías de reconocimiento de voz. En algunos casos, como en el dictado de recetas médicas y diagnósticos o el dictado de textos legales, se usan corpus especiales para incrementar la precisión del sistema.
* Control por comandos: Los sistemas de reconocimiento de habla diseñados para dar órdenes a un computador (p.e. "Abrir Firefox", "cerrar ventana") se llaman Control por comandos. Estos sistemas reconocen un vocabulario muy reducido, lo que incrementa su rendimiento.
* Telefonía: Algunos sistemas PBX permiten a los usuarios ejecutar comandos mediante el habla, en lugar de pulsar tonos. En muchos casos se pide al usuario que diga un número para navegar un menú.
* Sistemas portátiles: Los sistemas portátiles de tamaño reducido, como los relojes o los teléfonos móviles, tienen unas restricciones muy concretas de tamaño y forma, así que el habla es una solución natural para introducir datos en estos dispositivos.
* Sistemas diseñados para discapacitados: Los sistemas de reconocimiento de voz pueden ser útiles para personas con discapacidades que les impidan teclear con fluidez, así como para personas con problemas auditivos, que pueden usarlos para obtener texto escrito a partir de habla. Esto permitiría, por ejemplo, que los aquejados de sordera pudieran recibir llamadas telefónicas.

1. Funcionamiento

Un sistema de reconocimiento de voz está formado por varias capas o modelos. El primero de ellos es el modelo acústico, que permite a la tecnología identificar si el sonido procede de una llamada de móvil, de un teléfono IP o cualquier otro medio. Determinar el canal de comunicación es importante para establecer el grado de distorsión que puede experimentar el mensaje.

El segundo modelo es el modelo lingüístico y se trata del idioma. Pero no es tan sencillo como indicar al sistema que reconozca castellano, francés o mandarín. Un software puede estar perfectamente preparado para portugués, pero en ocasiones la gente suele hacer algunas pequeñas modificaciones, por ejemplo, el portugués que se habla en Madeira, que puede tener algún giro especial con el portugués tradicional. No sólo es preciso entender la lengua sino los distintos acentos con que se habla e incluso entender las formas de expresarse, que pueden ser diferentes en cada hablante.

Ésta es otra capa más y la última de ellas: el modelo semántico. Con él se consigue que un sistema de reconocimiento de voz entienda la forma de hablar de la gente, cómo se construyen las frases y cómo puede variar esta construcción, dependiendo de la región, de la cultura y de todas las influencias personales de cada cual. Por último, funciona un motor estadístico, que recoge la frase una vez transcrita a texto y realiza una búsqueda en la base de datos con estos términos.

Esta consulta tiene que adivinar si la frase dicha, con el nivel de distorsión correspondiente si se habla desde un móvil, el acento y la forma de decirlo están pidiendo una determinada acción. La precisión del software al final se reduce a aumentar la base de datos para alcanzar todas las combinaciones posibles en un idioma. Es más, en cada nuevo entorno que se implementa hay que construir un nuevo modelo semántico.

1. Identificación de timbres de voz

Dentro del reconocimiento de voz otra de las partes más destacadas es la biometría. Se trata de comprobar la identidad de una persona mediante la combinación única de patrones que tiene su voz. Para esto lo primero es obtener la huella vocal, que se puede sacar haciendo una captura dinámica mientras la persona habla, aunque también se puede establecer una frase como contraseña y que estas palabras precisas sean las que se comparen para la autentificación.

Esta huella registra las características de la voz, como pueden ser el timbre, el agudo, la edad o si es masculina o femenina. Pero también determina cuál es el canal por el que se está hablando, de manera que establece el posible grado de distorsión para ser capaz de reconocer al usuario a través de otro canal. Qué ocurriría si la voz es una grabación. El sistema es capaz de distinguir si es una grabación o si es una concatenación de palabras creada por una herramienta

1. Programas de reconocimiento de voz

Los sistemas de reconocimiento de voz pueden clasificarse según los siguientes criterios:

* Entrenabilidad: determina si el sistema necesita un entrenamiento previo antes de empezar a usarse.
* Dependencia del hablante: determina si el sistema debe entrenarse para cada usuario o es independiente del hablante.
* Continuidad: determina si el sistema puede reconocer habla continua o el usuario debe hacer pausas entre palabra y palabra.
* Robustez: determina si el sistema está diseñado para usarse con señales poco ruidosas o, por el contrario, puede funcionar aceptablemente en condiciones ruidosas, ya sea ruido de fondo, ruido procedente del canal o la presencia de voces de otras personas.
* Tamaño del dominio: determina si el sistema está diseñado para reconocer lenguaje de un dominio reducido (unos cientos de palabras p. e. reservas de vuelos o peticiones de información meteorológica) o extenso (miles de palabras).

1. Conclusiones

En la actualidad, la tecnología de reconocimiento de voz se ha convertido en algo habitual en nuestra vida diaria. Combinando nuestro método de comunicación más natural y antiguo, el habla, con los últimos avances tecnológicos, estamos, cada vez más, acostumbrándonos a realizar nuestras tareas cotidianas mediante comandos de voz y reconocimiento.

Los beneficios que trae el reconocimiento de voz para la cotidianidad son:

* Mejora la fiabilidad

Los sistemas de reconocimiento de voz actuales son tan fiables que están siendo ampliamente utilizados en sectores como la sanidad, el sector judicial, la industria de la seguridad o el sector militar. En este sentido, cada vez es más habitual ver a un médico dictar sus notas para convertirlas en documentos digitales o a abogados utilizar software de voz para registrar la información de sus clientes.

* Ahorra tiempo

Aquí es donde la tecnología de reconocimiento de voz puede ser especialmente útil, ya que nos permite dictar, en lugar de escribir, nuestro trabajo. Esto puede ser realmente valioso si tenemos poco tiempo, ya que simplemente tenemos que expresar nuestros pensamientos y dejar que el dispositivo haga el trabajo trascribiendo lo que decimos.

* Aumenta la productividad en el trabajo

Las notas de voz permiten producir una gran cantidad de información en un período relativamente corto. De esta forma, solo hablando de forma natural al micrófono y dejando que el software haga el resto, podemos fácilmente plasmar nuestros pensamientos en el papel. Esto nos permitirá tener más tiempo para la edición, redacción y revisión.

* Evita errores

Además de ser más rápida a la hora de completar las tareas, la tecnología de reconocimiento de voz es cada vez más precisa cuando se trata de vocabulario y ortografía. Más aún, existe la posibilidad de agregar nuestras propias palabras a la base de datos para ayudar a que el software nos entienda mejor.