Reconocimiento de audio

Soft Computing

Autor: Manuel Felipe Valencia Ceballos

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

Correo-e: felipe.valencia@utp.edu.co

***Resumen*— El presente documento se plantea la voz como aspecto único de las personas y como se implementaría esta en una aplicación donde se reconozca la voz e identifique quien es la persona que esta interactuando.**

***Palabras clave— Voz, personas, audio, reconocimiento.***

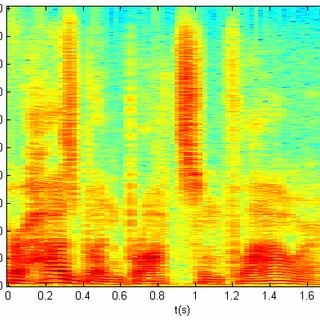
***Abstract*— This document presents the voice as a unique aspect of people and how it would be implemented in an application where the voice is recognized and identifies who is the person who is interacting.**

***Key Word*— Voice, people, audio, recognition.**

Toda persona posee aspectos particulares por el cual las personas las pueden identificar y recordar con facilidad, ya sea por su aspecto físico, por sus ojos, por su estilo. La voz también es un aspecto particular de las personas que las hace únicos. La voz es el sonido producido por un ser humano haciendo uso de sus cuerdas vocales para hablar, cantar, reírse, gritar, chillar, etc. Y al hacerlo, se produce una frecuencia que oscila entre alrededor de 60 a 7000 Hz. Así que las voces pueden ser medidas y vibrar a cierta frecuencia.

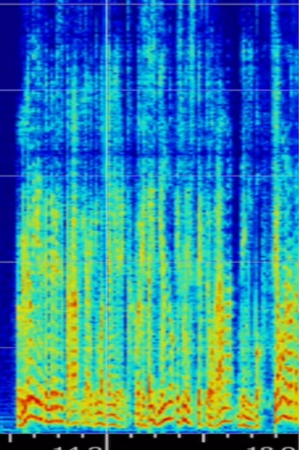
Al igual que la huella digital que hace únicas las personas, también se puede contar con un ID de la voz. Utilizar la huella digital como ID es la manera más eficiente que se ha desarrollado en los últimos años para reconocer personas, pero ¿Por qué no utilizar la voz como medio parar identificar las personas en diferentes áreas? ¿Cómo podría ser una posible aplicación que identifique a las personas por medio de su voz?

Hay que tener en cuenta que para reconocer se necesitan unos patrones a seguir. Por ejemplo, la huella dactilar es la impresión visible o moldeada que produce el contacto de las crestas papilares de un dedo de la mano (generalmente se usan el dedo pulgar o el dedo índice) sobre una superficie. O sea, gracias al relieve de las crestas papilares es que se puede marcar una huella y tener un patrón para reconocerla. ¿Pero cómo se podría esto con la voz? Cuando una persona habla, esta tiene cierto tono o melodía en su voz y estos. ¿Cómo identificar la intensidad de la voz de las personas?

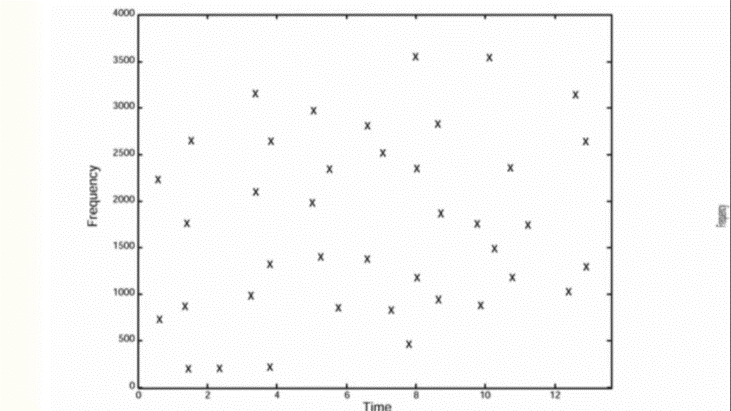


La imagen anterior es un espectrograma. Un espectrograma es una imagen que muestra la evolución temporal del espectro de una señal, el eje vertical va de las frecuencias más agudas a las más graves y el horizontal es el tiempo. En realidad, es una representación en tres dimensiones, temporal, frecuencial y amplitud de la distribución de energía de una señal.

Así que cuando una persona hable en el espectrograma se puede ver reflejado el resultado del tono de voz de la persona. ¿Esto para que pudiese servir en la construcción de la aplicación? Cuando una persona habla en su tono normal, va a tener unos puntos graves que serían las zonas más rojas en el espectrograma, así que se estos puntos rojos podrían ser marcados como los picos para identificar la voz de la persona.



Con la imagen anterior es el resultado de convertir una voz en una imagen, en un conjunto de datos. Con esto se puede hacer un análisis de frecuencia-tiempo para poder hacer match con la voz que esté interactuando. ¿Cómo? Las marcas se muestran en la siguiente imagen son puntos cables para identificarla. Este punto contiene una frecuencia en un determinado tiempo, y por cada punto se compara las frecuencias que tiene a su alrededor.



Todos y cada uno de estos datos estarían guardados en una base de datos con una capacidad muy grande, para así poder procesar cada una de las frecuencias y buscar como hacer match. De manera mejor explicada, si en el segundo dos (2) se registra una frecuencia de 300Hz, en este punto se puede basar para seguir haciendo su búsqueda, prosiguiendo con las frecuencias y tiempos vecinos a ver cual de estas coincide para poder hacer un match. Si se sabe que en ese punto hay una frecuencia de 300Hz y en el siguiente punto hay una frecuencia de 270Hz, con estos datos se busca en la base de datos justo en que momento y frecuencia se registra para avanzar en un posible match. Así con los puntos que estén alrededor de las frecuencias de 300Hz hasta que se haga un análisis completo de los puntos y se pueda hacer un match de manera efectiva.

Pensar en una solución completa para el diseño de reconocimiento de voz de una persona es demasiado compleja por diferentes aspectos a tener en cuenta, sin embargo, esta es una solución rápido que podría ser base para una pequeña solución para este planteamiento.

Para tener este este sistema completo se entraría a hablar del campo de la Inteligencia Artificial, para que este aprenda con la información que va percibiendo y reconocer con mayor eficiencia la voz de la persona que está interactuando e identificarla con facilidad.