Reconocimiento de locutores (Biometría de voz).

Speaker recognition (Voice biometrics).

Autor 1: Esteban Sánchez López
Universidad Tecnológica de Pereira
Correo-e: esteban.sanchez@utp.edu.co

***Resumen*— El documento desarrolla el tema del reconocimiento de locutores, también conocido como biométrica de voz, siendo esta una de las muchas ramas de la inteligencia artificial, durante el desarrollo de este documento se describirá en detalle en que consiste esta área, teniendo en cuenta áreas similares que interfieren o poseen un funcionamiento similar.**
Además, se hará énfasis en cada uno de los pasos/componentes del proceso de reconocimiento de locutores que hacen posible el correcto funcionamiento del proceso.
Se destacarán algunos de los campos de acción en los cuales se puede o ya se están aplicando de manera satisfactoria los procesos de biométrica de voz.

Palabras clave: Biométrica, inteligencia artificial, locutores.

***Abstract*— The document develops the topic of speaker recognition, also known as voice biometrics, this being one of the many branches of artificial intelligence, during the development of this document it will be described in detail what this area consists of, taking into account similar areas. that interfere or have a similar function.**
In addition, emphasis will be placed on each of the steps / components of the speaker recognition process that make the correct operation of the process possible.
Some of the fields of action in which voice biometric processes can or are already being successfully applied will be highlighted.

***Key Word* —Biometrics, artificial intelligence, announcers.**

I. INTRODUCCIÓN

El reconocimiento de locutores pertenece a la rama de la inteligencia artificial y consiste en la identificación automática de una persona a través de su voz. El hecho de poder distinguir un locutor de otro está relacionado mayoritariamente con las características fisiológicas y los hábitos lingüísticos de cada uno de ellos. El reconocimiento conlleva un procesado de audio que permite extraer este conjunto de rasgos inherentes al locutor y la posterior búsqueda de posibles coincidencias mediante un proceso de reconocimiento de patrones.

Conceptos básicos:
Antes de comprender el funcionamiento de estos sistemas es necesario tener claras algunas declaraciones y conceptos que nos permitirán entender más claramente el desarrollo del reconocimiento de locutores.

1. NATURALEZA DE LA VOZ.

En primer lugar y como es lógico por el objetivo del sistema y la función que desempeñará, lo primero que debemos entender en un sistema de reconocimiento de voz, es comprender la naturaleza misma de esta.
Lo primero que se puede observar al analizar la voz es que estas ondas no son estacionarias, es decir, si se toma una voz previamente grabada y se analiza su espectro de sonido, se observará como la onda poseerá ciertas características como pueden ser amplitudes diferentes a lo largo de la onda (esto incluso si se pronuncia la misma frase), pues por naturaleza la voz posee variaciones al ser generado por un sistema orgánico.

Ahora bien, a pesar de que ya se dijo que existen variaciones, esto depende también de que tan grande sea el periodo de tiempo que se analice, pues cuando se analizan ondas de periodos de tiempo del orden de milisegundos, se puede apreciar una tendencia de la onda a ser periódica, por lo que de este apartado se puede concluir que dependiendo de la cantidad de tiempo de la onda analizada (ya sea una onda del orden de segundos o de milisegundos) se obtendrán características especificas por lo que es un componente a tener en cuenta.

Figura 1. Locucion de 5 segundos.

2. PARAMETRIZACION DE LA VOZ.

La parametrización es definida como el proceso mediante el cual se extraen los parámetros característicos de las voces. Como se mencionó en el punto anterior, la voz analizada en diferentes extensiones, para obtener la información mas precisa de cada voz, y las características que se atañen a esta, lo mejor es tomar pequeñas muestras (al orden de los

Fecha de Recepción: (Letra Times New Roman de 8 puntos)
Fecha de Aceptación: Dejar en blanco

milisegundos), porque aquí las características que distinguen a cada voz.

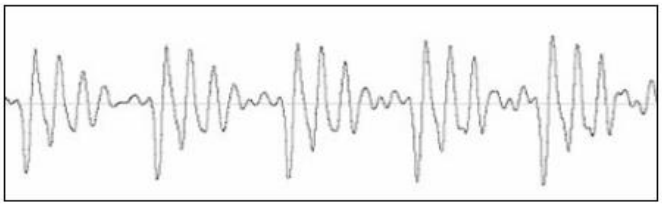

Figura 2. Locución de una vocal de 80 ms.

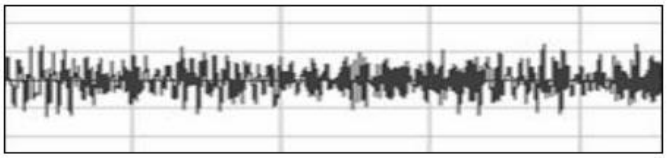

Figura 3. Tramo sordo de apariencia ruidosa.

Además, ya hablando un poco mas allá de la estructura de la voz en general, también hay que tener en cuenta que cada locutor posee características individuales, que principalmente son estas las que permitirán el reconocimiento y distinción entre los distintos locutores. Algunas de las características individuales son:

- Timbre de voz: Que indica dependiendo de la frecuencia de esa voz que tan grave o aguda es.
- Uso de los sonidos: Muchos locutores tienden a hacer u uso de sonidos a la hora de hablar que los caracterizan.
- Forma de entonar: Además del que se pronuncia, también es importante fijarse en el cómo, pues la entonación de las personas puede varias por diferentes factores.

Teniendo en cuenta estas características, mas adelante en el documento se hablará de la extracción de características donde se explicará como se agrupan estas en diferentes niveles.

II. ARQUITECTURA DEL SISTEMA

Un sistema de reconocimiento de locutor tiene que realizar ciertas tareas para que su funcionamiento sea correcto, las definiciones de estas son:

1) Adquisición de datos. La adquisición de datos es esencial tanto para la parte de entrenamiento como para la de testeo. De todas las pruebas de voz se almacenan los espectros característicos que se digitaliza y se almacena con los valores que acompañan la voz, como el nombre del locutor.

2) Extracción de características. Una vez digitalizado, el audio se procesa para extraer el listado de características elegidas, las cuales se llaman descriptores de audio. Estos descriptores contienen las características acústicas de la señal que utilizará el clasificador para compararlos con el listado almacenado en la base de datos. Las características para analizar pueden ser diversas, pero se suelen utilizar los descriptores de audio de bajo nivel debido a la naturaleza de la fuente. Estos descriptores presentan un bajo nivel de abstracción y se limitan a describir características espectrales, paramétricas y temporales de la señal de audio.

Las diferentes características se suelen agrupar en diferentes niveles, que es bastante útil al momento de analizar, estos niveles corresponden a:

• Fonético: utilización de Fonético: utilización de diferentes sonidos, diferentes sonidos, pronunciación, etc. pronunciación, etc.

• Prosódico: entonación particular, variación Prosódico: entonación particular, variación de energía, pausas entre frases o palabras, etc.

• Espectral: configuración (resonancia) del Espectral: configuración (resonancia) del tracto vocal, tracto vocal, coarticulación, nasalidad, etc. articulación, nasalidad, etc.

Para poder asociar las características de los descriptores a los archivos de audio correspondientes se utilizan los metadatos, datos sobre datos. Uno de los estándares utilizados para esta tarea es el estándar MPEG-7, el cual permite la gestión de estos metadatos, facilitando así el acceso a esta información en el momento de la búsqueda.

Esta tarea se puede dividir en módulos mas pequeños que cumplen funciones mas específicas, esto es útil para que al momento del desarrollo en caso de errores sean más fáciles de solucionar al no afectar tantas partes del sistema.

El módulo llamado Preprocesador Acústico, convertirá la señal acústica de entrada en una serie de vectores de características que extraigan de forma eficiente la información de locutor presente en la señal de voz. Opcionalmente se podrán incluir funciones para dotar de mayor robustez acústica el sistema.

El módulo de patrones/referencias dispondrá de patrones o referencias correspondientes a los distintos locutores conocidos por el sistema (usuarios) y obtenidos en la fase de entrenamiento.

3) Clasificación. El módulo clasificador tiene acceso tanto a la parte de entrenamiento como a la de testeo. Este módulo hace de puente entre ambas partes encargándose de comparar los vectores de características a buscar con los vectores de los modelos de locutor que contiene la base de datos. Su tarea computacional consiste en encontrar coincidencias y como resultado extrae una serie de probabilidades de los locutores en la base de datos susceptibles de ser el buscado. La decisión puede ser diferente dependiendo de la configuración del sistema.

Dentro de esta tarea existe un modulo mas especifico, denominado El módulo de cálculo de similitudes, una vez obtenidos los vectores de características correspondientes a la señal de voz de entrada, y teniendo disponibles los modelos o

patrones correspondientes a los distintos locutores, calculará el parecido o similitud entre la realización acústica de entrada y cualquiera de los modelos conocidos por el reconocedor.

4)Toma de decisiones: a partir de los valores de similitud obtenidos, deberá tomar una decisión acerca de la identidad del locutor que ha generado la locución de entrada.

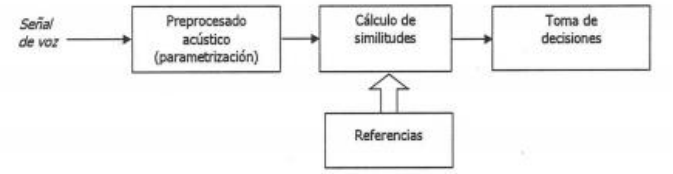

Figura 4. Estructura del sistema.

III. TIPOS DE SISTEMAS

Una vez conocidas las diferentes tareas y procesos que debe realizar los sistemas de reconocimiento de locución, se procederá a definir los tipos de sistemas que hay dependiendo de las funciones para las que se enfocan estos.

Actualmente los sistemas de reconocimiento de locutores se dividen en 2 tipos descritos a continuación:

Sistema cerrado: Un sistema cerrado da por supuesto que el locutor que se quiere identificar se encuentra ya almacenado en la base de datos. El locutor con más probabilidades a la salida del clasificador, que comparte más características con el locutor a buscar, será la salida resultante del sistema.

Sistema abierto: Un sistema abierto es más complejo, ya que el locutor que se quiere identificar no está necesariamente en la base de datos. El clasificador debe tener en cuenta no sólo la más alta probabilidad, sino que también debe establecer si la semejanza es suficiente para dar un positivo. Si las probabilidades de un modelo de locutor se consideran suficientes como para suponer una coincidencia se presenta al candidato como resultado de la búsqueda, en caso contrario la salida es "locutor desconocido".

IV. ENFOQUE

En este apartado vamos a describir brevemente en qué consisten las diferentes Tareas de Reconocimiento que puede realizar un Sistema Automático de Reconocimiento de Locutor. Dependiendo cual sea su objetivo podremos clasificar las tareas en dos grandes grupos que describiremos a continuación. Se trata de las tareas de identificación y las tareas de verificación

IDENTIFICACIÓN

El objetivo de una identificación de locutores es el de clasificar una señal de voz, cuyo origen no conocemos, como perteneciente a uno de entre un conjunto de N posibles locutores.
Dentro de estos sistemas, debemos diferencias dos posibles casos:
Identificación en conjunto cerrado: en este caso, el resultado del proceso es una asignación de identidad a uno de los locutores modelados por el sistema, y conocidos como «usuarios». Existen, por tanto, N posibles salidas posibles

Identificación en conjunto abierto: Aquí debemos considerar una posibilidad adicional a las N del caso anterior, y es que el locutor que pretende ser identificado no pertenezca al grupo de usuarios, con lo que el sistema de identificación debería contemplar la posibilidad de no clasificar la locución de entrada como perteneciente a las N posibles.

VERIFICACIÓN

En la verificación de locutores, en contraposición a la identificación se reciben dos entradas. Una de ellas es la señal de voz a verificar y la otra es una solicitud de identidad, que puede ser realizada de diversas formas. De este modo, las dos únicas salidas o decisiones del sistema son la aceptación o rechazo de la hipótesis de que ambas locuciones pertenezcan a la misma persona. La decisión de aceptar o rechazar la locución de entrada como correspondiente al locutor solicitado dependerá de si el valor del parecido o probabilidad obtenido supera o no un determinado umbral de decisión.

TIPO DE ENTRADA

Finalmente, Se pueden dividir también los diferentes sistemas dependiendo de una característica bastante importante y es el método mediante el cual reconoce y obtiene sus respuestas, esta característica la entrada que proporciona el locutor, y se divide en si la entrada es un texto predefinido o no.

Los sistemas dependientes del texto utilizan la misma palabra o frase tanto en la parte de entrenamiento como en la de testeo. Estas palabras suelen ser contraseñas privadas en aplicaciones de seguridad.
Los sistemas independientes del texto no se basan en ninguna palabra o frase en concreto y no necesitan ningún tipo de cooperación por parte del locutor a buscar, pues con la voz ya es suficiente. Estos sistemas se utilizan a menudo en campos de investigación forense o judicial, para identificar a locutores o verificar alguna identidad.

V. APLICACIONES

El desarrollo de tecnologías encargadas de reconocer automáticamente a una persona mediante su voz ha experimentado un creciente interés en los últimos años debido a sus múltiples aplicaciones.

1. Las notas de pie de página deberán estar en la página donde se citan. Letra Times New Roman de 8 puntos

CAMPO	EJEMPLOS
Control de acceso	• Acceso a instalaciones físicas • Acceso a un ordenador
Transacciones de autenticación	• Comercio electrónico • Transacciones bancarias
Servicio personalizado	• Aplicaciones de domótica
Gestión de audio	• Indexación automática de contenidos de audio
Refuerzo de la ley	• Comprobación de que se cumple la libertad condicional
Forense	• Identificación de personas a través de grabaciones para validar pruebas

VI. DESAFIOS ACTUALES

• Variabilidad de la voz entre sesiones
• Sigue siendo muy problemático en condiciones extremas.
• Degradación del rendimiento con poco material de voz
• Locuciones cortas (típicamente de prueba)
• Desajuste de base de datos
El sistema se entrena con datos en condiciones muy diferentes a la de funcionamiento real (ruido, estilo de habla, reverberación, etc.) habla, reverberación, etc.)

VII.CONCLUSIONES

El área de reconocimiento de locutores presenta características bastante destacables al momento de realizar procesos, demuestra la necesidad de examinar un gran número de aspectos como, por ejemplo, la voz humana. Cabe destacar que como ya se mencionó anteriormente, el área cuenta con una variedad enorme de áreas de aplicación, en donde se podrían facilitar y mejorar procesos enormemente.

REFERENCIAS

Documentos presentados en conferencias (No publicadas aún):

[1] http://garciaargos.com/descargas/apuntes/posgrado/Primer-Semestre/Reconocimiento-Biometrico/2008_Master_UAM_Locutor_v4.pdf
[2] https://es.wikipedia.org/wiki/Reconocimiento_de_locutores#Arquitectura_del_sistema
[3] https://www.ub.edu/journalofexperimentalphonetics/pdf-articles/XVII-17.pdf