

Computación Blanda

Reconocimiento Automático del Locutor

Autor: Alejandra López Ocampo

IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia
Correo-e: Alejandra.lopez1@utp.edu.co

Resumen— El presente documento busca abordar el problema de identificación de personas públicas por medio de una grabación de su voz. En este se hace referencia al tema de reconocimiento automático del locutor como herramienta para la verificación de identidad a partir de esta característica biométrica. Asimismo, se habla de como deben ser empleados los modelos de extracción de datos y entrenamiento para generar un algoritmo que pueda aprender a distinguir o identificar la voz.

Palabras clave— biometría, sistemas, extracción de características, entrenamiento, memoria, comportamiento.

Abstract— This paper seeks to address the problem of the identification of public persons through a recording of their voice. In this document, reference is made to the topic of automatic speaker recognition as a tool for identity verification based on this biometric characteristic. Likewise, it talks about how data extraction and training models should be used to generate an algorithm that can learn to distinguish or identify the voice.

Key Word— biometrics, systems, extraction of characteristics, training, memory, behavior..

I. INTRODUCCIÓN

La biometría es un campo que a pesar de que sus estudios pueden resultar recientes se practica desde el principio de los tiempos. De manera inconsciente hacemos uso de la biometría, un ejemplo muy claro de esto es al momento de contestar el teléfono, debido a que intentamos comprobar que la voz que escuchamos al otro lado de la línea coincide con alguna de las que tenemos almacenadas en nuestra memoria.

Las tecnologías que están diseñadas para reconocer de una manera automática y rápida la voz de las personas, han experimentado un creciente interés en los últimos años. Lo anterior se debe a que esto trae consigo múltiples aplicaciones que podrían ser bastante útiles para distintos campos.

Es importante destacar que se le comenzó a dar más importancia a este campo gracias a que muchos investigadores en criminología vieron en la biometría una herramienta efectiva para medir características físicas de un individuo.

La manera en la que reconocemos la voz es un proceso bastante complejo dentro del cual influyen muchas variables. Esto debido a que la señal de la voz en si misma depende de características tanto emocionales como físicas como el estado de ánimo, de salud, los hábitos lingüísticos y la posición y dimensiones de las cuerdas vocales. La diferencia puede llegar a ser tan grande y varía tanto de individuo a individuo que incluso puede ser usado como una característica biométrica.

SISTEMAS DE RECONOCIMIENTO BIOMÉTRICO

Podemos definir un equipo biométrico como “Aquel que tiene capacidades para medir, codificar, comparar, almacenar, transmitir y/o reconocer alguna característica propia de una persona, con un determinado grado de precisión y confiabilidad.” (Sistemas biométricos). Teniendo esto en cuenta, se puede pensar en que es la manera ideal para identificar a un individuo evitando otros métodos que sean susceptibles a la falsificación, a la transferencia y a las técnicas de decodificación.

Los sistemas de reconocimiento biométrico funcionan a partir de alguna característica de una persona que pueda identificarlo de manera particular. Los métodos con los cuales se pueden obtener estas características las clasifican en dos como lo son las intrusivas como las pruebas de sangre y las no intrusivas como la palma de la mano, la forma de hablar y caminar.

Estos por lo general constan con componentes tanto de hardware como de software para poder realizar el proceso de reconocimiento. El concepto biometría a partir de su etimología ya que proviene de las palabras bio (vida) y metría (medida), por lo tanto, se puede decir que todo equipo biométrico mide e identifica alguna característica propia de la persona.

CLASIFICACIÓN DE SISTEMAS BIOMÉTRICOS

Con la evolución del campo de la biometría se encontró la necesidad de clasificarla en dos tipos. La biometría estática y la biometría dinámica.

La biometría estática por su lado es la que se ha usado de manera convencional, por ende, es la más antigua, esta consiste en la medición y tratamiento de características estancadas, es decir que no pueden cambiar como lo son las huellas dactilares o la morfología del rostro. Por otro lado, con el pasar de los años y el desarrollo de nuevas tecnologías se ha abierto paso a lo que hoy conocemos como biometría dinámica, que gracias a un proceso de recogida y tratamiento de los datos adecuado puede usar, por ejemplo, la manera en la que la persona camina para su identificación.

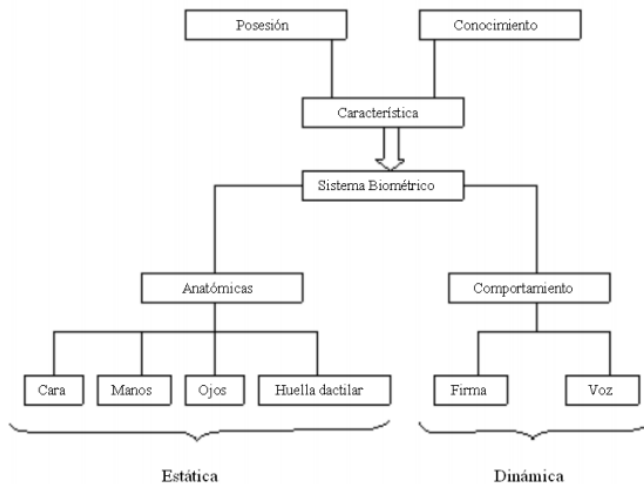


Figura 1. Clasificación de los sistemas biométricos.

II. AUTENTIFICACIÓN DE LA VOZ

Dentro del campo de identificación del locutor que ahora es considerada como una “biométrica del comportamiento” (Lo cual puede ser evidenciado en la figura anterior) es importante entrar a hacer ciertas distinciones con fines prácticos. Por lo tanto, dentro del campo del reconocimiento del habla se han concertado dos grandes campos:

- Verificación del hablante
- Identificación del hablante

En el campo de la verificación del hablante se tiene como objetivo procesar la muestra de voz y compararla con las que ya se tenían almacenadas para determinar si pertenece a la persona quien se dice, es decir, busca una identidad “afirmada”

Por otro lado, en la identificación del hablante la identidad de la persona que suministra la muestra de voz es totalmente desconocida y su objetivo es como su nombre lo indica, es el poder identificarla. Pero en ambos casos se necesitan muestras de voces previas para el reconocimiento.

En los sistemas de reconocimiento de voz no se intenta entender o reconocer lo que el usuario está diciendo, ya que su prioridad es identificar una serie de sonidos y características particulares. “La representación matemática de una señal de voz, se centra en la descomposición de la señal como una fuente pasando a través de un filtro variable en el tiempo, partiendo de un modelo que aproxima el funcionamiento del órgano humano productor de la voz”

La extracción de estos patrones acústicos sacados de la señal de voz forma lo que es conocido como un “patrón de voz” o una “huella de voz”

En cuanto al problema del reconocimiento del locutor se intenta dar una solución por medio de la clasificación de audios gracias a un modelo automatizado, el cual fue entrenado previamente con una base de datos de autos suministrada previamente. Dentro de esta se busca el audio del usuario con mayor coincidencia. Por consiguiente, estos sistemas cuentan con cierta estructura que permite su adecuada implementación.

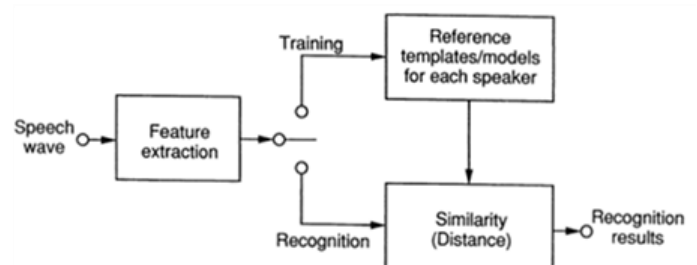


Figura 2. Estructura de un sistema de identificación del locutor.

Por otra parte, es importante agregar que el proceso también tiene ciertas variantes, ya este tipo de verificación puede hacerse tomando como base un texto específico que tendría que ser replicado por la persona a la cual se le va a realizar el reconocimiento o, por otro lado, hacerla independiente del contexto, donde el contenido del audio en sí no es relevante.

Ahora se explicará de una manera breve algunas partes del proceso de identificación del hablante:

- Adquisición: Se recibe una grabación en un formato preestablecido que cumpla con unos mínimos estándares, ya que cualquier situación irregular puede afectar los resultados.
- Preprocesamiento: Se adecua el audio y se remueve el ruido lo cual consiste en frecuencias no importantes en la voz (Las menores a 30Hz y mayores a 4000Hz)
- Extracción de características: El resultado de esta es un fotograma de señal de voz donde es representada de manera matemática
- Clasificación y toma de decisiones: Haciendo uso de las redes neuronales, se hace la clasificación en base a su proceso de aprendizaje automático

REDES NEURONALES ARTIFICIALES

Estas intentan simular la estructura y funciones del cerebro humano. Pueden definirse como un sistema de procesamiento de información, en este caso en específico, señales. Las RNA están conformadas por nodos, los cuales están interconectados por medio de uniones directas que permiten el procesamiento.

Las RNA son sistemas de clasificación efectivos ya que su implementación se suele realizar con datos incompletos y ruidosos, con lo cual son perfectas para este tipo de problemas que tiene dichas características.

En el documento “Identificación Automática del Hablante mediante Redes Neuronales” se usaron perceptrones multicapa para la clasificación y un algoritmo de retropropagación con el cual se buscó un número óptimo de capas donde se mejorará dicho proceso:

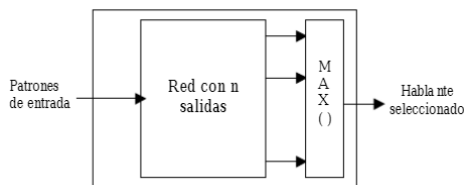


Figura 3. Estructura de clasificador tradicional

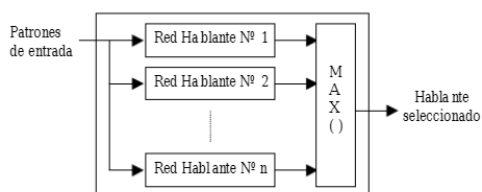


Figura 3. Estructura de clasificador propuesto por Torres Humberto M., Rufiner Hugo L.

III. CONCLUSIONES

El estudio de la biometría nos permite avanzar en el desarrollo de sistemas infalibles en cuanto al reconocimiento de personas se refiere, proporcionando nuevas técnicas y equipos para que esto sea posible.

IV. REFERENCIAS

Referencias en la Web:

- [1] http://www.ai.org.mx/ai/archivos/ingresos/nolazco/rabajo_final.pdf
- [2] <https://planetachatbot.com/clasificador-autom%C3%A1tico-de-audios-a-partir-de-la-identificaci%C3%B3n-del-locutor-abfec66e9bef>
- [3] <https://www.ub.edu/journalofexperimentalphonetics/pdf-articles/XVII-17.pdf>
- [4] <http://147.83.50.136/talp/index.php/es/investigacion/lineas-de-investigacion/rec-locutor>
- [5] <http://recursostic.educacion.es/observatorio/web/fr/software/software-general/689-reconocimiento-y-sintesis-de-voz>
- [6] <https://planetachatbot.com/clasificador-autom%C3%A1tico-de-audios-a-partir-de-la-identificaci%C3%B3n-del-locutor-abfec66e9bef>
- [7] <https://docs.microsoft.com/es-es/learn/modules/recognize-voices-with-speaker-recognition/2-overview-of-speaker-recognition>
- [8] https://www.dsi.uclm.es/personal/MiguelFGraciani/mikicurri/Docencia/Bioinformatica/web_BIO/Documentacion/Trabajos/Biometria/Trabajo%20Biometria.pdf
- [9]

<https://repositoriotec.tec.ac.cr/bitstream/handle/2238/5652/reconocimiento%20de%20voz.pdf?sequence=1&isAllowed=y>

[10]

https://www.researchgate.net/publication/265964745_Identificacion_Automatica_del_Hablante_mediante_Neural_Networks