Reconocimiento de locutores.

speaker recognition.

Juan David Osorio Ortiz
Universidad tecnológica de Pereira, Pereira, Colombia
Correo-e: juandavid.osorio1@utp.edu.co

***Resumen*— el reconocimiento de locutores es el proceso de extracción automática de información relativa a la identidad de la persona a partir de muestras de voz. El proceso tiene dos etapas básicas: entrenamiento (recolección de muestras de voz de las personas a ser identificadas) y reconocimiento (comparación de las muestras del locutor desconocido con los datos de entrenamiento, y toma de decisión).**

***Palabras clave*— reconocimiento de locutores, ia, computación blanda, reconocimiento de voz, voz, automatización, ondas sonoras, seguridad informática.**

***Abstract*— speaker recognition is the process of automatic extraction of information regarding the identity of the person from voice samples. The process has two basic stages: training (collection of voice samples from the people to be identified) and recognition (comparison of the unknown speaker samples with the training data, and decision making).**

***Key Word* — speaker recognition, ai, soft computing, speech recognition, voice, automation, sound waves, computer security**

I. INTRODUCCIÓN

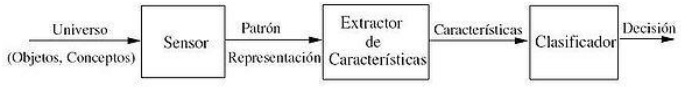
En este documento se va a realizar una introducción sobre el reconocimiento de locutores en el cual se mencionará la arquitectura de un sistema de reconocimiento de locutores, las aplicaciones y algunas otras definiciones importantes sobre este tema.

II. CONTENIDO

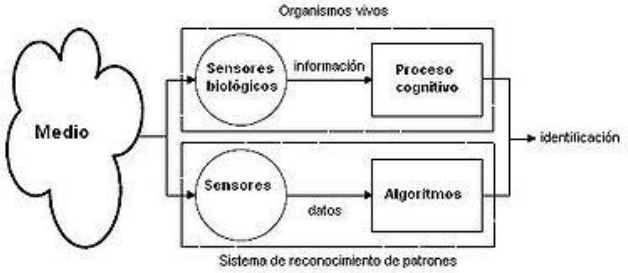
¿Qué es el reconocimiento de locutores?

Es una de las ramas de la inteligencia artificial que consiste en la identificación automática de una persona a través de su voz. El hecho de que se pueda distinguir un locutor de otro está relacionado mayoritariamente con las características fisiológicas y los hábitos lingüísticos de cada uno de ellos. Este reconocimiento conlleva un procesado de audio que permite extraer este conjunto de rasgos pertenecientes a la voz del locutor y posibles coincidencias mediante un proceso de reconocimiento de patrones.

Para poder que se de un reconocimiento de algún patrón se deben seguir una serie de procesos:



- Adquisición de datos: Para este paso de normal se usan sensores por los cuales se pueden captar magnitudes físicas o químicas y transformarlas a magnitudes eléctricas para que estas puedan ser procesadas en el siguiente paso.



- Extracción de características: En este proceso se generan características que pueden ser usadas en el proceso de clasificación de datos. En ocasiones se precede de un preprocesado de la señal para corregir posibles deficiencias en los datos debido a errores del sensor.

- Toma de decisiones: Una vez analizados los datos y clasificados según lo encontrado en las características recibidas el programa busca similitudes con muestras que ya se habían ingresado para dar la respuesta más acertada.

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

¿Verificación vs Identificación?

Estos dos campos mencionados son los mas utilizados dentro del área de reconocimiento de locutores. Si el locutor afirma tener una determinada identidad y el sistema es el encargado de corroborar esta afirmación, se está realizando verificación de locutores. Si por otro lado el sistema solo recibe características de una voz y debe determinar su identidad.

En la verificación de locutores el sistema de reconocimiento verifica si las características extraídas de la voz del locutor corresponden con la identidad que el afirma tener. El resultado de este proceso siempre es una verificación binaria en la que el resultado es o éxito en caso de que se haya verificado o fracaso en caso de que las características no coincidan. Una de las aplicaciones mas frecuentes de la verificación es en sistemas de seguridad.

En un sistema de identificación se suelen recibir una o más muestras de voz, y estas son contrastadas con una base de datos con voces cuyas identidades ya son conocidas para el sistema. Luego el sistema retorna una puntuación de semejanza entre las muestras y la información de la base de datos. Siendo mayor entre mas similitud se tenga entre la muestra y el registro de la base de datos.

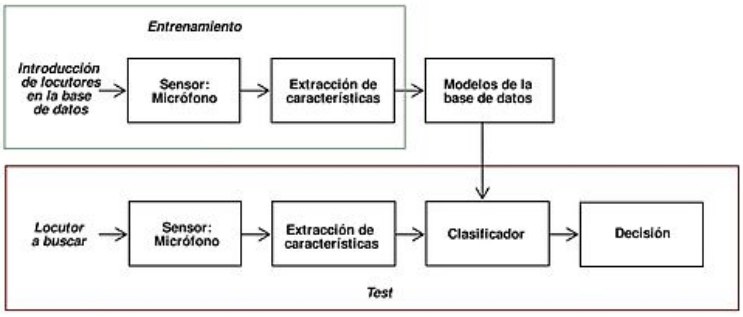
Arquitectura del sistema.

Un sistema de reconocimiento de locutor esta conformado por dos secciones:

- Entrenamiento.
- Test.

Entrenamiento: tiene la finalidad de registrar locutores mediante un micrófono para extraer sus características y guardarlas en la base de datos.

Test: se centra en registrar a un locutor y extraer las características para poder compararlas con las que se encuentran almacenadas en la base de datos. Finalmente, después de obtener posibles coincidencias, el sistema presenta al locutor susceptible de ser el buscado.



Arquitectura de un sistema locutor

Adquisición de datos en reconocimiento de locutores:

La adquisición de datos es esencial tanto para la parte de entrenamiento como para la de test. Para poder introducir locutores al sistema es necesario un transductor acústico-eléctrico, ya que la voz se propaga en forma de ondas y para poder extraer características es necesario transformar la presión sonora en una señal eléctrica y así poder proceder a su digitalización.

Micrófonos (Transductor acústico)

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.

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.

```
<?xml version="1.0" encoding="iso-8859-1" ?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001
Mpeg7-001.xsd">
<Description xsi:type="ContentEntityType">
    <MultmediaContent xsi:type="ImageType">
        <image name="ph00000.jpg">
            <MediaLocator>
              <MediaURi>
              fichier:c://testimag//ph00000.jpg
              </MediaURi>
            </MediaLocator>
            <VisualDescriptor
            xsi:type="DominantColorType">
              <SpatialCoherency>31</SpatialCoherency>
              <Value>
                 <Percentage>31</Percentage>
                 <Index>27 25 22</Index>
                 <ColorVariance>0 0 0</ColorVariance>
              </Value>
            </VisualDescriptor>
        </image>
    </MultmediaContent>
  </Description>
</Mpeg7>
```

MPEG-7 (file example)

Clasificación.

El módulo clasificador tiene acceso tanto a la parte de entrenamiento como a la de test. 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.

Tipos de sistemas de Reconocimiento de locutores.

- 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".

Aplicaciones del reconocimiento de locutores.

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. Algunos de estos campos de aplicación son los siguientes:

- Control de acceso (Seguridad).
- Transacciones de autenticación (Seguridad).
- Servicios personalizados (Domótica)
- Gestión de audio (Multimedia).
- Refuerzo de la ley (Seguridad).
- Forense (Seguridad).

Como se puede ver actualmente el reconocimiento de locutores tiene una gran importancia en el campo de la seguridad informática debido a que actualmente la importancia de los datos y de la seguridad en todos los aspectos a tomado demasiada relevancia en los últimos años.

Métodos de reconocimiento de locutores.

Enfoque de ajuste de Plantillas (Template Matching): Para distinguir entre distintos locutores se compara un promedio de los vectores característicos obtenidos en la etapa de reconocimiento usando un segmento relativamente largo de señal de voz, con una colección de promedios obtenidos durante la etapa de entrenamiento.

Modelado Probabilístico de Locutores: Los vectores característicos se consideran variables aleatorias caracterizados por una función de distribución de probabilidad en vez de por valores promedios (media y covarianza). La clasificación es basada en medidas de verosimilitud en vez de en medidas de distorsión o distancia entre vectores de reconocimiento y de referencia.

Asumiendo que los locutores tienen una distribución conocida, con funciones de densidad de probabilidad continuas, entonces la probabilidad de que un vector característico x haya sido generado por el i-ésimo locutor es usando la Regla de Bayes.

III. CONCLUSIONES

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

1. El reconocimiento de locutores es una rama que tiene un crecimiento exponencial y aunque actualmente la mayoría de las aplicaciones de esta es en seguridad, en unos años se va a estar aplicando de muchas maneras comerciales.
2. A pesar de que actualmente no es uno de los mejores métodos de seguridad por que puede tener muchas vulnerabilidades debido a hardware o posibles faltas de seguridad por su sencillez para poder romper su seguridad en unos años va a ser el pilar de seguridad en los sistemas informáticos si se sigue investigando en esta área como se a hecho hasta el momento.

REFERENCIAS

1. https://es.wikipedia.org/wiki/Reconocimiento_de_locutores

2. https://www.fceia.unr.edu.ar/prodivoz/speaker_verification.pdf

3. https://es.wikipedia.org/wiki/Reconocimiento_de_patrones