Reconocimiento de voz, a través de una entrada de audio.

Speech recognition via audio input.

Autor: Daniel Felipe Quintero

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

Correo-e: [Daniel\_feli10@utp.edu.co](mailto:Daniel_feli10@utp.edu.co)

***Resumen*— La Inteligencia Artificial ha experimentado un gran crecimiento durante los últimos años, gracias al aumento de la capacidad de cómputo y la aparición de herramientas que permiten realizar cálculos complejos de una forma mas eficiente. En esta investigación, se realiza una búsqueda de como poder resolver el problema de que un sistema inteligente sea capaz de realizar la identificación de una persona a través de un audio, actualmente el uso de los fonemas tiene implícita varias dificultades debido a que la identificación de las fronteras entre ellos, por lo regular es difícil de encontrar en representaciones acústicas de voz. Para construir estos sistemas se utiliza la información conocida sobre el tracto vocal de un individuo y las propiedades físicas del sonido**

***Palabras clave—* sistemas, redes, inteligencia artificial, software, computación, investigación, industria, genético, aprendizaje.**

***Abstract*— Artificial Intelligence has experienced great growth in recent years, thanks to increased computing capacity and the emergence of tools that allow complex calculations to be performed more efficiently. In this research, a search is carried out on how to solve the problem of an intelligent system being able to perform the identification of a person through an audio, currently the use of phonemes has implicit several difficulties because the identification of the boundaries between them, is usually difficult to find in acoustic representations of voice. To build these systems, known information about an individual's vocal tract and the physical properties of sound is used.**

***Key Word*— systems, networks, artificial intelligence, software, computing, research, industry, genetic, learning.**

1. INTRODUCCIÓN

El presente trabajo plantea una alternativa a la forma en la que el reconocimiento de voz se ha estado implementando desde hace ya bastante tiempo. Las grandes empresas del sector tecnológico han apostado recientemente por el desarrollo de asistentes virtuales, pudiendo encontrar disponibles soluciones de Apple (Siri), Microsoft (Cortana), Amazon (Alexa), entre otras.

Un sistema de reconocimiento de voz es aquel sistema automático que es capaz de obtener elementos de medición (muestras), las cuales permiten denotar su comportamiento e implementar procesos de tratamiento de la señar, enfocados al reconocimiento.

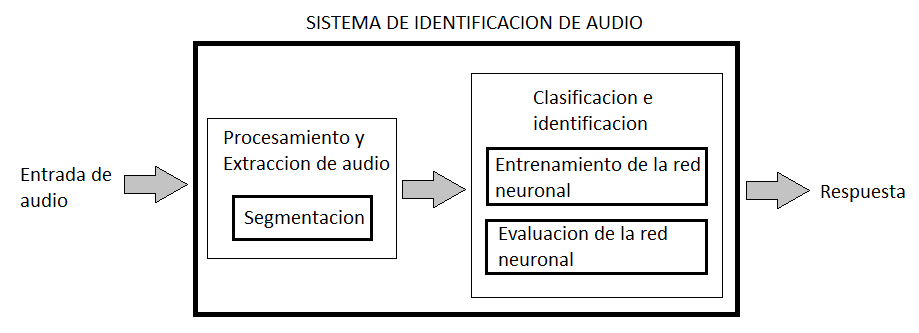
La voz humana es definida por la Real Academia de la Lengua como “Sonido que el aire expelido de los pulmones produce al salir de la laringe, haciendo que vibren las cuerdas vocales” o “Calidad, timbre o intensidad de este sonido” [RAE]. El habla consiste en combinar las unidades fónicas (fonemas) que la voz humana es capaz de generar para formar una lengua, estos símbolos u hondas no son idénticos de un individuo a otro,

Actualmente, los sistemas pueden dividirse en dos grandes ramas de estudio, la autentificación biométrica, que trata de verificar la identidad del hablante, y el estudio bajo entrenamiento, que esta orientado a reconocer la voz de un individuo entre un catalogo de voces ya estudiadas y analizadas previamente, este ultimo es el que nos centraremos en esta investigación.

1. DESARROLLO

II.1. Principio para solución del problema:

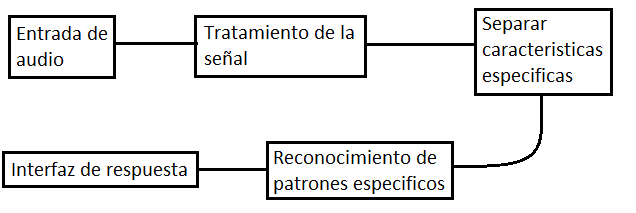
Para desarrollar el sistema de reconocimiento, se llevaron a cabo los procesos mostrados en el esquema de la figura 1. En el diagrama se incluye las fases de obtención de señales a través del sistema embebido y un computador, el proceso de las mismas, utilizando segmentación y algoritmia de extracción de características, como los coeficientes de las frecuencias.



II.2 Área de trabajo:

Para realizar una mejor explicación, vamos a realizar un diagrama un poco simplificado, pero con mucho aporte intelectual, este modelo de reconocimiento de voz, mostrando en la figura 2, y sugerimos unas partes importantes a considerar a la hora de llevar esta solución. Las principales áreas de trabajo desde diseño y especificación, encontramos:

* Transformación de la señal
* Reconocimiento de patrones
* Búsqueda de estilos de habla
* Reconocimiento de voz
* Bases de datos de los audios ya guardados



Podemos observar en que básicamente consiste en analizar una señal de voz, primero extrayendo características de ella, luego pasando por una serie de patrones, para así después el sistema poder aprenderla y reconocerla mediante una red neuronal.

II.3 Reconocimiento de voz:

En este primer estado de la propuesta que estamos tratando de diseñar, el programa de reconocimiento de voz analizará mediante la metodología de redes neuronales, una selección de los audios que se ingresa y un procesado de la voz que entra al sistema, con el objetivo de extraer solo la información acústica relevante de los audios, esta dinámica está dividida en dos partes, aprendizaje y prueba:

* Aprendizaje:
  + Adquisición de Datos
  + Pretratamiento de la señal
  + Extracción de características
  + Entrenamiento de la red neuronal
* Pruebas:
  + Adquisición de Datos
  + Pretratamiento de la señal
  + Extracción de características
  + Clasificación

II.3.1 Adquisición de datos:

El primer paso es la adquisición de señales de voz (en este caso audios) que servirán para construir nuestra base de datos inicial, esta se usara tanto en la parte de entrenamiento como en la parte de pruebas.

II.3.2 Pretratamiento de la señal:

Después de la adquisición de datos se observa que cada muestra ingresada contiene datos que resultan en parte inútiles, por lo cual el siguiente paso es la eliminación de tales segmentos, lo cual se logra tomando como referencia los niveles iniciales y finales de la muestra para obtener un promedio.

Seleccionados los datos inútiles, se procede a la normalización de las muestras, ya que la amplitud de cada una difiere, y para obtener mejores resultados, es deseable que todas las muestras estén en la misma escala, en este caso, de -1 a 1.

II.3.3 Extracción de características:

La extracción de parámetros o características obtiene vectores de un determinado objeto que modelan su patrón. Según los estudios de la voz, se dice que esta puede simularse como un filtro vocal del cual puede representarse un espectro de las componentes frecuenciales de la voz y extraer información representativa

II.3.4 Entrenamiento de la Red Neuronal:

Para proceder con el entrenamiento de la red neuronal, se puede ejecutar con la utilización de un vector de n entradas (entre más mayor sea el n, mejor) se obtiene un vector de pesos ajustados que se utilizan para la etapa de pruebas.

Cada audio de entrenamiento proporciona una matriz de coeficientes que se irán concatenando hasta formar la matriz de entrenamiento, que será la que contenga los coeficientes de todos los audios, cabe recordar, que estos audios ya pasaron por los pasos de Extracción de características y pretratamiento, para que este, en esté paso ya este listo para que el sistema lo guarde en su base de datos.

REFERENCIAS

Referencias en la Web:

[1]

<http://www.scielo.org.mx/pdf/cys/v9n3/v9n3a7.pdf>

<http://oa.upm.es/51541/1/TFG_JUAN_MANUEL_VICENTE_CABERO.pdf>

<https://e-archivo.uc3m.es/bitstream/handle/10016/26403/TFG_Raquel_Donoso_Garcia_Del_Castillo.pdf?sequence=1&isAllowed=y>