Computación Blanda

Soft Computing

Autor: Juan Manuel Sanchez Pareja
IS&C, Universidad Tecnológica de Pereira, Pereira, Colombia
Correo-e: juan manuel.sanchez@utp.edu.co

***Resumen*— El reconocimiento de voz es un tipo de inteligencia artificial que trata de establecer una comunicación entre el hombre y los ordenadores o los dispositivos inteligentes, a través del lenguaje humano. El mecanismo de voz hace que el procesador sea capaz de descifrar la información que contiene la voz humana. Los primeros pasos en este aspecto comenzaron en los años 50, pero ha sido en la última década cuando se ha hecho un avance gigantesco en el reconocimiento del lenguaje natural.**

***Palabras clave*— Reconocimiento de voz, reconocimiento de sílabas, sistemas expertos, procesamiento de voz**

***Abstract*— Voice recognition is a type of artificial intelligence that tries to establish communication between man and computers or smart devices, through human language. The voice mechanism makes the processor capable of deciphering the information contained in the human voice. The first steps in this regard began in the 1950s, but it has been in the last decade that a gigantic advance has been made in the recognition of natural language**

I. INTRODUCCIÓN

Durante los últimos años han surgido tecnologías de interfaces humano-computadora que combinan varias tecnologías del lenguaje, analizando la forma en el cual el paradigma de la sílaba responde a tal labor como dentro del español. Un Sistema de Reconocimiento Automático de Habla (SARAH) es aquel sistema automático capaz de gestionar la señal emitida por un individuo. Dicha señal ha sido pasada por un proceso de digitalización para obtener elementos de medición (muestras), las cuales permiten denotar su comportamiento e implementar procesos de tratamiento de señal, enfocados al reconocimiento.

Bajo este esquema, la señal de voz se ve inmersa en dos bloques importantes: entrenamiento y reconocimiento. Dicho entrenamiento, es una de las etapas más críticas dentro de estos sistemas y gran parte del éxito de un sistema de reconocimiento de voz recae en esta etapa. Como un referente esencial de lo anterior, el presente trabajo presenta la incrustación de un bloque destinado al refuerzo de la obtención de los datos a ser procesados que hace uso de un Sistema Basado en Conocimiento (SBC al cual se le denominará también Sistema Experto), capaz de realizar la clasificación de la señal de entrada en unidades silábicas, por medio de la aplicación de un conjunto de reglas lingüísticas que prevalecen en el español. La razón por la cual se pensó en un Sistema Basado en Conocimiento es debido a que en el español en contraparte del inglés por ejemplo, la forma en la que se escriben los textos y la que se lee no dista mucho de ser semejante. Esto es debido a que el español es altamente dependiente del contexto y de la prosodia. Los elementos anteriores justifican la aplicación del experto en esta parte del sistema.

II. Contenido

II.1 Historia

La historia del reconocimiento de voz empezó en el año de 1870. Alexander Graham Bell quiso desarrollar un dispositivo que capaz de proporcionar la palabra visible para la gente que no escuchara. Bell no tuvo éxito creando este dispositivo, sin embargo, el esfuerzo de esta investigación condujo al desarrollo del teléfono. Más tarde, en los años 30 Tihamer Nemes científico húngaro quiso patentar el desarrollo de una máquina para la transcripción automática de la voz. La petición de Nemes fue negada y a este proyecto lo llamaron poco realista.

Fue hasta 1950, 80 años después del intento de Bell, cuando se hizo el primer esfuerzo para crear la primera máquina de reconocimiento de voz. La investigación fue llevada a los

laboratorios de AT&T. El sistema tuvo que ser entrenado para reconocer el discurso de cada locutor individualmente, pero una vez especializada la máquina tenía una exactitud de un 99 por ciento de reconocimiento

Fue hasta 1950, 80 años después del intento de Bell, cuando se hizo el primer esfuerzo para crear la primera máquina de reconocimiento de voz. La investigación fue llevada a los laboratorios de AT&T. El sistema tuvo que ser entrenado para reconocer el discurso de cada locutor individualmente, pero una vez especializada la máquina tenía una exactitud de un 99 por ciento de reconocimiento

Durante los 60 's, La mayoría de los investigadores reconoce que era un proceso mucho más intrincado y sutil de lo que habían anticipado.
Flujo discreto de habla (con espacios / pausa entre palabras)Vocabulario pequeño (menor o igual a 50 palabras).Estos sistemas empiezan a incorporar técnicas de normalización del tiempo (minimizar diferencia en velocidad del habla).Además, ya no buscaban una exactitud perfecta en el reconocimiento.

Durante los 80's el reconocimiento de voz se favoreció por tres factores: el crecimiento de computadoras personales, el apoyo de ARPA y los costos reducidos de aplicaciones comerciales. El mayor interés durante este periodo de tiempo era el desarrollo de vocabularios grandes. En 1985 un vocabulario de 100 palabras era considerado grande. Sin embargo, en 1986 hubo uno de 20,000 palabras. También durante esta época hubo grandes avances tecnológicos, ya que se cambió del enfoque basado en reconocimiento de patrones a métodos de modelado probabilísticos, como los Modelos Ocultos de Markov

II.2 Las características acústicas de los sonidos del habla

La onda sonora(tono puro)

La onda sonora producida en la fonación es el resultado del paso del aire por la glotis en la emisión de una serie de sucesivas bocanadas de aire al ritmo de abertura y cierre de los pliegues vocales.

Para la producción de la onda sonora las moléculas de aire deben entrar en vibración, lo que se consigue por su paso a través de los pliegues vocales.

Movimiento vibratorio de una molecula de aire

El ciclo de un movimiento vibratorio:

- Fuerza que desplaza la molécula;
- Alejamiento de la posición inicial de reposo;
- Vuelta a la posición de partida debido a la elasticidad;
- Desplazamiento hacia el lado contrario por inercia:
- Vuelta a la posición inicial.

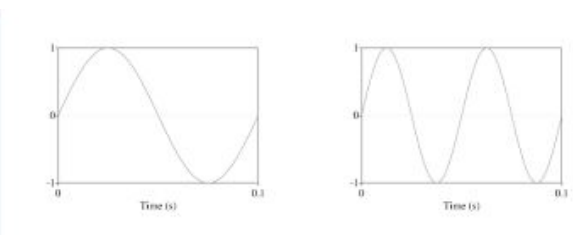
Amplitud del movimiento vibratorio

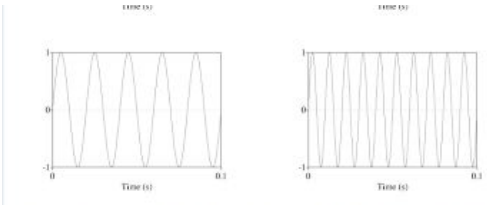
- Distancia entre la posición de reposo y el punto de máximo desplazamiento.
- Depende de la fuerza aplicada inicialmente y de la resistencia que ofrezca el medio en el que se produce la vibración.
- La amplitud se cuantifica en unidades de presión sonora desde el punto de vista físico o en unidades de intensidad (Decibelios, dB) desde el punto de vista perceptivo.

Parámetros acústicos que caracterizan una onda sonora

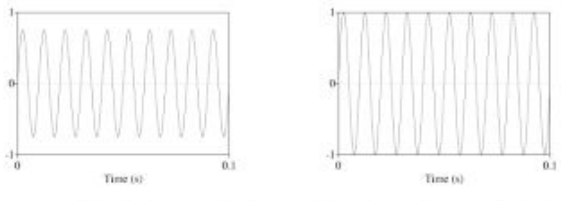
La onda sonora es el resultado de la vibración de las moléculas del aire y puede definirse en función de:

- su amplitud
- su frecuencia
- el tiempo durante el cual se lleva a cabo el movimiento





Ondas sonoras periódicas simples con la misma amplitud (1) y con diferente frecuencia (10 Hz, 20 Hz, 50 Hz, 100 Hz).



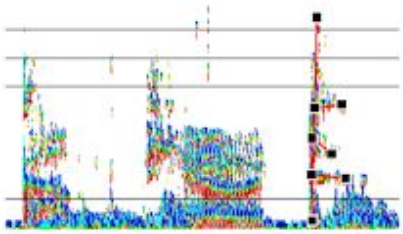
Ondas sonoras periódicas simples con la misma frecuencia (100 Hz) y con diferente amplitud (0,10, 0.50, 0.75, 1).

II.3 Representación de la señal de voz

Los sonidos consisten en variaciones en la presión del aire a través del tiempo y a frecuencias que podemos escuchar. Una de las maneras para representar el sonido es a través de una onda (waveform)

Figura 1.1 Ejemplo de una señal de voz

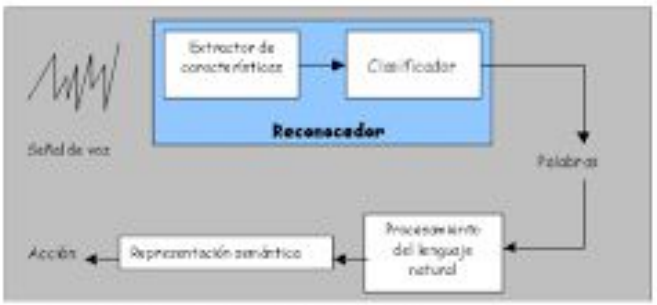
Una de las grandes ventajas de éste tipo de gráficas es que no ocupa mucho espacio en memoria. Y una desventaja es que no se describe explícitamente el contenido de la señal en términos de sus propiedades. Los espectrogramas contienen mayor información sobre los datos de la voz, son una transformación que muestran la distribución de los componentes de frecuencia de la señal



Las partes más oscuras, vistas en la figura, representan la concentración de energía y son denominadas formantes. Por otra parte es importante mencionar que la capacidad auditiva del ser humano varían en un rango de frecuencias de 20Hz a 20,000Hz. Los sonidos emitidos al hablar se encuentran de 100Hz a 15,000Hz en mujeres y en hombres de 400Hz a 15,000Hz. Aunque se cree que la mayoría de la información se concentra debajo de los 8,000Hz

I.1 Arquitectura de un sistema de reconocimiento de voz

Para entender el funcionamiento de un sistema de reconocimiento de voz es necesario conocer sus principales componentes: el extractor de características y el clasificador. Cuando se recibe la señal de voz, ésta pasa por un reconocedor el cual da como resultado la palabra que reconoce. Después hay un procesamiento del lenguaje natural, una representación semántica y finalmente se realiza una acción. La arquitectura para los sistemas de reconocimiento de voz se muestra a continuación:



Como se mencionó anteriormente, en la arquitectura de un sistema de reconocimiento de voz se cuenta con dos procesos importantes en la fase de reconocimiento, estos son los siguientes

Extracción de características: Los pasos a realizar en este módulo son los siguientes

- La señal se divide en una colección de segmentos.
- Se aplica alguna técnica de procesamiento de señales para obtener una representación de las características acústicas más distintivas de segmento
- con base a las características obtenidas, se construye un conjunto de vectores que constituyen la entrada al siguiente módulo

Clasificador probabilístico: Los pasos a realizar en este módulo son los siguientes:

- Se crea un modelo probabilístico basado en redes neuronales como modelos ocultos de Markov, etc
- Con las probabilidades obtenidas se realiza una búsqueda para encontrar la secuencia de segmentos con mayor probabilidad de ser reconocidos.

I.2

REFERENCIAS

Referencias en la Web:

[1] https://invoxmedical.com/noticias/la-inteligencia-artificial-en-el-reconocimiento-de-voz/
[2] https://www.timetoast.com/timelines/procesamiento-digital-de-la-voz
[3] http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html
[4] http://www.cochlea.eu/es/sonido/representacion#:~:text=Representación%20tridimensional%3A%20el%20sonograma%20o,analizar%20la%20señal%20de%20voz.