

Señales De Tiempo Discreto

Proyecto Final J.A.R.V.I.S

Camilo Mejía Martínez¹, Daniel Pineda Suárez², Julian Oviedo Pasuy³

¹ mejiacamiloe@javeriana.edu.co ² dapineda@javeriana.edu.co ³ julian.oviedo@javeriana.edu.co

Facultad de Ingeniería, Departamento de Electrónica
Pontificia Universidad Javeriana
2019-I

Fecha de Entrega : Mayo 31 de 2019

Resumen—El diseño y desarrollo de una aplicación de reconocimiento de voz para el asistente robótico de Iron Man presentado a continuación, surgió como una solución al requerimiento de una aplicación de los conocimientos adquiridos en la asignatura de Señales de Tiempo Discreto, una vía de comunicación alternativa hombre-máquina. Esta tecnología constituye una herramienta sencilla que reconoce comandos de voz, cuyo alcance esencial es el ser aplicado en áreas como robótica y domótica, relacionando el reconocimiento de patrones de voz con posibles funciones de control de sistemas inteligentes. Se desarrolló un sistema de reconocimiento de comandos de voz sobre una plataforma de simulación (MATLAB y AppDesigner), basado en la extracción de parámetros de la voz su comparación en frecuencia. Se empleó el análisis de la densidad espectral cruzada y su comparación por medio de la correlación entre el comando ingresado y un comando base grabado previamente. El sistema es dependiente del locutor dado que la fuente fue la voz perteneciente a una única persona y se llevaron a cabo experimentos para un sistema de conjunto cerrado, señales de las bases de datos propias y para un sistema de conjunto abierto o señales grabadas en tiempo real. Se determinó la mejor combinación de parámetros con la cual se logró una tasa de aciertos del 80 %.

Index Terms—Transformada rápida de Fourier, reconocimiento de comandos de voz, producto interno.

I. INTRODUCCIÓN

Siri, Cortana, Google Now, Amazon Echo, son algunos de los asistentes de voz que se encuentran compitiendo por convertirse en la mejor inteligencia artificial del reconocimiento de voz.

La inteligencia artificial cada vez se perfecciona más, podríamos decir que los avances más cercanos en inteligencia artificial ya se encuentran alrededor de nosotros.

El habla o la comunicación verbal es una de las formas más naturales de comunicación para el ser humano.

Respondiendo al requerimiento de implementar un sistema en el que se apliquen los conocimientos adquiridos en Señales de Tiempo Discreto se implementa un sistema de reconocimiento de comandos o patrones de voz.

La fase de Concepción y Diseño responden a preguntas cómo: ¿cuáles son las técnicas más apropiadas para desarrollar un sistema de reconocimiento de voz?, y qué otras utilidades o aplicaciones se pueden derivar a partir de esta tecnología?. Las principales ventajas obtenidas al implementar un control por voz en un asistente es la aplicación en control de sistemas

inteligentes, en áreas como domótica, robótica, y en el sector automotriz.

El presente trabajo se desarrolló como proyecto final, profundizando principalmente en los conceptos de Procesamiento y Análisis de Señales Discretas mediante la Transformada de Fourier, obteniendo como resultado una aplicación desarrollada con AppDesigner en Matlab. Se evaluaron distintas técnicas de análisis y extracción de características de voz, como análisis cepstral (Cepstrum), wavelets, coeficientes de predicción lineal (LPC), sin embargo, estas técnicas quedan fuera del alcance del curso y por ende de la aplicación de estos en el desarrollo de este proyecto. Se puso énfasis en la evaluación de los criterios de generación de diccionarios o bases de comandos, como base de comparación.

Por último, se estudiaron las distintas alternativas aplicadas para el bloque del sistema reconocedor, como Modelos Ocultos de Markov (HMM) y redes neuronales artificiales, sin embargo el método usado para la implementación del proyecto es el cálculo de distancia métrica realizando su procesamiento mediante la correlación cruzada, con la se obtuvo buenos resultados como se presentarán más adelante.

La descripción de la construcción del módulo de reconocimiento y control por voz que se desarrolla en el presente artículo como respuesta a la primera cuestión.

II. REVISIÓN BIBLIOGRÁFICA

II-A. Transformada discreta de Fourier(TDF)

La transformada discreta de Fourier, DFT por sus siglas en inglés discrete Fourier transform, de una señal de longitud L y de N muestras se define como:

$$TDF_N\{x[n]\} = X\{k\} = \sum_{n=0}^{N-1} x[n]e^{-j\Omega_k n} \quad (1)$$

donde $\Omega_k = \frac{2\pi k}{N}$ corresponde a las N frecuencias evaluadas tomado de [1]

II-B. Transformada rápida de Fourier(FFT)

La transformada rápida de Fourier, FFT por sus siglas en inglés Fast Fourier Transform, permite realizar la Transformada discreta de Fourier de forma mas eficiente ya que para la computación de la DFT se deben realizar N^2 operaciones

mientras que con la FFT tan solo se necesitan $N \log_{10}(N)$. Existen muchos tipos de FFT, uno de los mas famosos es el algoritmo de Cooley_Tukey el cual consiste en el diezrado en base 2 para así reducir calcular la DFT para segmentos mas cortos de la señal. tomado de [1]

II-C. Espacios vectoriales

Un espacio vectorial o lineal es una colección de objetos, denominados vectores, que pueden ser sumados y multiplicados por un escalar. Todo espacio vectorial normado es también un espacio métrico, con métrica definida como $p(v,w) = ||v-w||$.

$$p(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2)$$

tomado de [1]

II-D. Espacio vectorial $L1$

se define como:

$$\{f : Z \rightarrow C : \sum_{n=-\infty}^{\infty} |f| < \infty\} \quad (3)$$

lo conforman todas las señales reales y complejas sumables en magnitud. tomado de [1]

II-E. Espacio vectorial $L2$

Este conjunto se define como:

$$\{f : Z \rightarrow C : \sum_{n=-\infty}^{\infty} |f|^2 < \infty\} \quad (4)$$

lo conforman todas las señales reales y complejas sumables en magnitud al cuadrado a demás por su definición esta conformado por todas las señales de energía y se puede comprobar que es un espacio vectorial euclidiano. tomado de [1]

II-F. Formantes del habla

Estos son los picos de intensidad en el espectro de frecuencia donde hay concentración de energía es la máxima amplitud de onda que se da en una determinada frecuencia. Estos permiten reconocer los diferentes sonidos del habla humana, cada sonido del habla tiene una marca característica de formantes por lo cual sirven para los sistemas de reconocimiento de voz. tomado de [2]

II-G. Frecuencias del habla

Las frecuencias conversacionales de la voz humana están entre 250 y 3000 Hz aunque las cuerdas vocales pueden modular la voz de 20 a 20000 Hz. Sin embargo se sabe que para transmitir la voz de manera entendible no es necesario transmitir toda esta banda frecuencia, los teléfonos transmiten en un rango aproximado de 400 a 4000 Hz, esto distorsiona la voz pero no lo suficiente como para no entender. tomado de [3]

II-H. Producto interno

El producto interno, denotado como $\langle \cdot, \cdot \rangle$, se define para dos elementos del espacio vectorial y asigna un elemento de un espacio vectorial V , con campo asociado F , de tal forma que $\forall v, w \in F$ y $\forall a \in F$ satisfaga:

$$\langle v, v \rangle \geq 0 \text{ y } \langle v, v \rangle = 0 \Leftrightarrow v = 0 \quad (5)$$

II-I. Micrófono

un micrófono es un elemento capaz de transformar las ondas sonoras en ondas eléctricas por medio de dos transductores, uno transforma las vibraciones de presión en vibraciones mecánicas, el otro recibe estas vibraciones mecánicas y las transforma en magnitudes eléctricas equivalentes. tomado de [4]

II-J. Matlab

Matlab abreviatura de MATrix LABoratory, es una plataforma poseedora de su propio lenguaje de programación la cual es usada principalmente por científicos e ingenieros en todo el mundo por sus usos posibles entre las cuales esta cálculos matriciales, desarrollo y ejecución de algoritmos, creación de interfaces de usuario y visualización de datos. Estos usos son usados en campos como procesamiento de imágenes y señales, comunicaciones, sistemas de control, diseño de redes inteligentes, robótica, y finanzas computacionales. tomado de [5]

II-K. .WAV

Los archivos de este tipo .WAV (Wave Audio File) contienen datos de audio digital que se pueden ver en variedad de reproductores de audio digital, estos archivos suelen ser mas grandes en comparación a otros formatos como el MP3 ya que su contenido de audio no esta comprimido. Los datos de estos archivos también se llaman formas de onda las cuales se pueden variar de velocidad siendo el valor por defecto de 44100 Hz. tomado de [7]

II-L. Correlación cruzada

En los procesos digitales muy comúnmente se debe interpretar el grado de interdependencia entre dos procesos o la similitud entre dos señales, entre sus campos de aplicación se encuentra la detección e identificación de señales. Se puede relacionar el espectro de una señal con la transformada de Fourier de su función correlación, cuando se aplica a dos señales se obtienen la función de densidad espectral cruzada, que se define como la transformada de Fourier de la función correlación cruzada. tomado de [8] y tomado de [9]

III. OBJETIVOS

III-A. General

- Implementar con la herramienta Matlab un sistema capaz de identificar un comando de voz, usando los conocimientos adquiridos en clase.

III-B. Específicos

- Generar un Base de Comandos de voz para la comparación.
- Implementar un sistema de reconocimiento de habla usando la comparación con la base de comandos, usando el método más apropiado.
- Crear una interfaz gráfica con AppDesigner para la interacción del usuario con el sistema implementado.

IV. METODOLOGÍA

El sistema está desarrollado principalmente en 4 fases, las cuales son: Creación y consolidación de una base de comandos, adquisición de señal y acondicionamiento de la voz (Digitalización), Procesamiento y Comparación por Correlación Cruzada de las señales, Reconocimiento del comando respectivo e implementación de la interfaz gráfica(aplicación desarrollada en AppDesigner para interacción con el sistema). Podemos tener en cuenta, la siguiente figura para entender el flujo del sistema.



Figura 1. Diagrama de bloques del sistema de comparación

Se consolida una base de datos con los distintos comandos a reconocer, estos archivos de audio se encuentran en formato ".wav", lo que permite tener una mejor calidad en el sonido puesto que no hay compresión de los datos, por ende mejor calidad en ellos.

Tanto para el audio adquirido como para la base de comandos se realiza la digitalización de estos con una Frecuencia de Muestreo de 44100 Hz para tener alta fidelidad .

Se procede a realizar el procesamiento mediante la correlación cruzada de los vectores generados al digitalizar las grabaciones de los comandos; normalizando el vector se obtiene la función densidad espectral cruzada (cross-spectrum), que se define como la Transformada de Fourier de la función correlación cruzada, cuando la función vale 1 para todas las frecuencias, se dice que $x(t)$ e $y(t)$ son totalmente coherentes. Cuando para alguna frecuencia w_0 vale cero, se dice que son incoherentes a esa frecuencia. Si ambas señales son estadísticamente independientes, la función de coherencia vale cero para todas las frecuencias.

Por último se reconoce el comando comparando las distancias y estimado el mínimo error entre ellas.

La aplicación implementada tendrá como salida para cada caso una reproducción que responderá ante el reconocimiento de un comando en específico, simulando de esta manera, un sistema inteligente, como lo es J.A.R.V.I.S.



Figura 2. Interfaz Gráfica del Sistema(Aplicación)

El protocolo a seguir para el uso de la aplicación consta de tres sencillos pasos:

Primero se deberá hacer clic en Grabar, con lo que podremos obtener la señal y almacenarla para su posterior procesamiento y análisis.

Luego procedemos a escucharla para saber si quedó parecida a la que esta almacenada en el diccionario.

Por último, se deberá hacer clic en detectar, que desencadenara su procesamiento, análisis y comparación para obtener como resultado el comando reconocido.

V. RESULTADOS Y DISCUSIONES

Los resultados mostrados a continuación corresponden a las pruebas aplicadas al sistema realizadas con éxito, se presentan las gráficas de los espectros de las señales tanto de la biblioteca o base de comandos como el de la señal de entrada, para tener un punto de comparación visual que permita reconocer patrones de amplitud en la frecuencia.

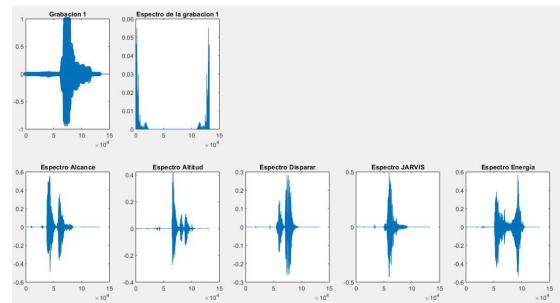


Figura 3. Comando Alcance

La figura presentada corresponde al reconocimiento del comando 'Alcance' para lo cual se gráfica su gráfica temporal y la magnitud de su Transformada Rápida de Fourier (FFT).

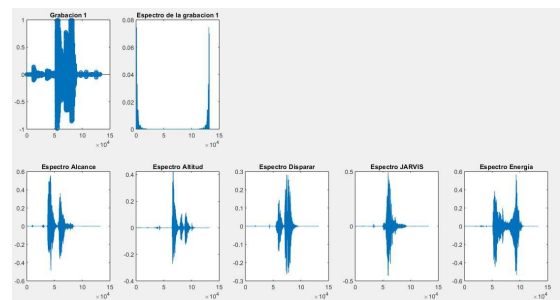


Figura 4. Comando Altitud

La figura presentada corresponde al reconocimiento del comando 'Altitud' para lo cual se gráfica su gráfica temporal y la magnitud de su Transformada Rápida de Fourier (FFT).

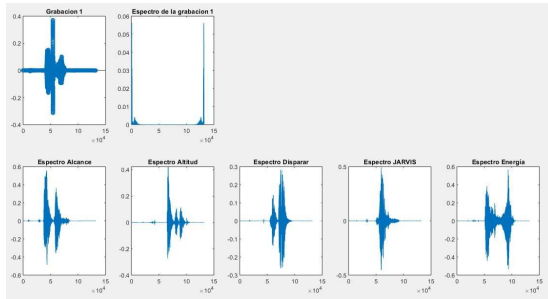


Figura 5. Comando Disparar

La figura presentada corresponde al reconocimiento del comando 'Disparar' para lo cual se gráfica su gráfica temporal y la magnitud de su Transformada Rápida de Fourier (FFT).

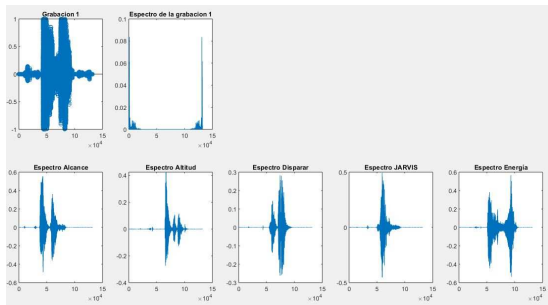


Figura 6. Comando Energía

La figura presentada corresponde al reconocimiento del comando 'Energía' para lo cual se gráfica su gráfica temporal y la magnitud de su Transformada Rápida de Fourier (FFT).

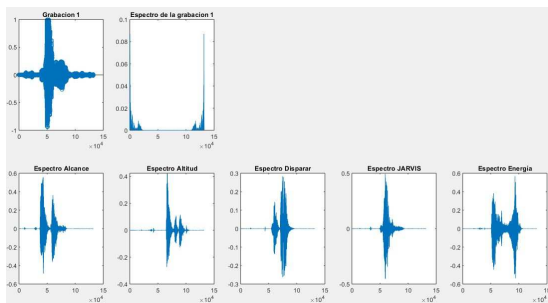


Figura 7. Comando Jarvis

La figura presentada corresponde al reconocimiento del comando 'Jarvis' para lo cual se gráfica su gráfica temporal y la magnitud de su Transformada Rápida de Fourier (FFT).



Figura 8. Comando Reconocido en Interfaz Gráfica del Sistema

En La figura anterior se puede observar el funcionamiento correcto de la aplicación y por ende del sistema, reconociendo de esta manera el comando 'Altitud'.

Se debe tener en cuenta al analizar el espectro de las señales que el ruido ambiente es mínimo, puesto que el diccionario se creó en un ambiente silencioso lo que limita el proyecto a ser probado en este tipo de ambiente.

La aplicación de un filtro para disminuir el ruido no fue eficaz en la reducción de ruido puesto que atenuaba el volumen de la señal y no aportaba al proceso de reconocimiento.

La tasa de aciertos se logra calculando el número de aciertos sobre el número de intentos, esta se cumple siempre que las condiciones ambientales antes mencionadas estén presentes.

Aunque el proyecto cuenta con limitaciones, podemos observar que obtuvimos una tasa de aciertos aceptable, lo que demuestra que el sistema es eficaz si se cumple con las condiciones necesarias para su óptimo funcionamiento.

VI. CONCLUSIONES

- El sistema implementado es de manera sencilla una herramienta que permite el reconocimiento de comandos de voz, podemos decir que es adaptativo, pues se pueden cambiar los comandos de voz de la biblioteca por los de otra persona y así realizando el mismo proceso poder reconocer en diferentes voces comandos distintos.
- Las limitaciones de la solución propuesta radican en el filtrado del ruido ambiente, como solución de pre-énfasis ante la adquisición de la señal, por lo que el alcance próximo o sugerencia es mejorar la robustez del sistema, para lograr que funcione correctamente en ambientes ruidosos o con otras voces de fondo, lo que implicaría un procesamiento de mayor nivel.
- El uso de los conceptos adquiridos en la asignatura de Señales Discretas fueron útiles para entender conceptos como correlación, producto interno, norma y espacios vectoriales, sin embargo se logró con lecturas relacionadas con el tema y asesoría sobre los procesos que implementamos un buen resultado acorde a los alcances del curso, fortaleciendo de esta manera nuestros conceptos y habilidades relacionadas con esta interesante área.

REFERENCIAS

- [1] J. QUIROGA, *Fundamentos de señales y sistemas*, primera edición, Bogotá D.C, Colombia, Pontificia universidad javeriana, febrero de 2018.
- [2] ANONIMO, *Transmision de la voz*, (online), Available: <http://elastixtech.com/fundamentos-de-telefonía/transmision-de-la-voz/>

- [3] ANONIMO ,1 de mayo de 2019, *Formantes*,(Online), Available: <http://musiki.org.ar/Formantes>
- [4] ANONIMO , *microfonos*,(Online), Available: <http://www.ehu.eus/acustica/espanol/electricidad/micres/micres.html>
- [5] MARGARET ROUSE , *Matlab*,abril de 2015,(Online), Available: <https://whatistechtarget.com/definition/MATLAB>
- [6] ANONIMO , *.WAV*,(Online), Available: <https://www.reviversoft.com/es/file-extensions/wav>
- [7] ANONIMO , *.WAV*,(Online), Available: <https://www.reviversoft.com/es/file-extensions/wav>
- [8] J.F GUERRERO MARTINEZ , *Estimación espectral*,2010-2011,(Online), Available: http://ocw.uv.es/ingenieria-y-arquitectura/1-5/ib_material/IB_T5_OCW.pdf
- [9] ANONIMO , *CORRELACIÓN CRUZADA Y CORRELACIÓN*,(Online), Available: <http://www.ehu.eus/Procesadodesenales/tema8/corre1.html>