



**UNIVERSIDAD MICHOACANA DE SAN
NICOLÁS DE HIDALGO**



**FACULTAD DE CIENCIAS
FÍSICO MATEMÁTICAS**

**Reconocimiento de voz por
medio de la transformada de
Fourier**

T E S I S

Que para obtener el título de:

LICENCIADO EN CIENCIAS FÍSICO MATEMÁTICAS

Presenta:

JOSÉ LUIS ROJAS CASTAÑEDA

Asesor:

Dr. MAURICIO ORTIZ GUTIÉRREZ

Co-asesor:

Dra. YENISEY DEL ROCÍO PONCE DE LEÓN VILLANUEVA

Morelia, Mich.

Diciembre 2021

Dedicatoria

ESTA TESIS ESTÁ DEDICADA A MIS PADRES JOSÉ LUIS Y CLAUDIA, A MIS HERMANOS LIZETH Y JUAN DIEGO, A MIS ABUELAS CARITINA Y MARIA ELIA, A MI BISABUELA ARCADIA LEDESMA QUE ESTÁN O ESTUVIERON CONMIGO DESDE PEQUEÑO. POR SUS ENSEÑANZAS, TODO SU APOYO Y AMOR.

GRACIAS.

Agradecimientos

Gracias a mis padres por tanto amor y enseñanzas a lo largo de este camino, me han enseñado a querer y trabajar duro por lo que quiero, mostrarme el camino, que cuando más perdido estaba ellos eran la luz que podía seguir para levantarme a pesar de las adversidades. Por enseñarme que con esfuerzo y dedicación cualquier cosa se puede lograr. A mis hermanos Juan Diego y Viridiana que han estado conmigo en todo momento, tanto felices como tristes, que a pesar de todo los quiero.

Gracias al Dr. Mauricio por darme su tiempo y orientarme en este trabajo, aguantarme en los momentos más difíciles y estar apoyándome en todo momento. Agradezco también al laboratorio de óptica por haberme dado un espacio para trabajar en las primeras etapas de este trabajo, por las herramientas y por contar un espacio con el que podía trabajar, siendo de mucha ayuda en los momentos en los que se decidió como experimentar.

A mis compañeros y mejores amigos, Leonardo, Marco Antonio, Diego, Andrea, Mariana, Luis Antonio, Luis, Alberto, Felipe, Pedro, Javier, Gustavo, Daniela y Mauricio gracias por todos los buenos momentos y por las risas que nunca faltaron, sin esos momentos todo habría sido más difícil. Muchas gracias. A mis compañeros de laboratorio Andrés, Francisca, Daniela, Andrea que con ellos siempre podía encontrar un momento de relajación y apoyo. Muchas gracias.

Agradecerle a mi novia Natalia por apoyarme en este trabajo, darme los ánimos en todo momento y tenerme paciencia durante este trabajo.

Gracias a mis abuelas Maria Elia y Caritina por estar conmigo desde pequeño apoyarme con su sabiduría e historias que me han guiado y he usado como ejemplo para mi día a día.

Resumen

El uso de la transformada de Fourier es muy amplio tomando en cuenta que es una herramienta muy poderosa e importante ya que con ella se pueden describir fenómenos físicos de óptica, resolver ecuaciones diferenciales y tiene aplicaciones en el área médica. La transformada de Fourier ha llegado a convertirse en un excelente instrumento de trabajo en nuestro día a día. Esto se puede ver en el uso de aplicaciones móviles como lo son Shazam entre otras. Tomando en cuenta esta idea, el objetivo a desarrollar en el presente trabajo, es comprender cómo funciona la transformada de Fourier para poder reconocer a una persona por medio de su voz cuando se tienen a un grupo de n personas.

Es importante entender primeramente la teoría del sonido y definir todas las propiedades físicas del mismo para tener una visión más clara del funcionamiento de los programas de identificación de voz.

Posteriormente una vez explicada la teoría del sonido se presenta la teoría de la transformada de Fourier, empezando por la definición, su variante en dos dimensiones y la demostración de ciertos teoremas que tienen un valor importante en la teoría del sonido.

Después se enfoca la descripción del problema, dando una descripción más a detalle de la transformada de Fourier. Como elemento importante dentro de esta investigación, se generó un código en Matlab para facilitar la imagen del espectro de voz de una persona.

Por último se muestran los resultados, conclusiones y observaciones de los datos recabados en la experimentación de este trabajo.

Palabras clave: Acústica, Sonido, Espectro de sonido, Convolución, Conjugado.

Abstract

The use of the Fourier transform is very broad taking into account that it is a very powerful and important tool since with it it is possible to describe physical phenomena of optics, solve differential equations and has applications in the medical area. The Fourier transform has become an excellent working tool in our daily lives. This can be seen in the use of mobile applications such as Shazam among others. Taking this idea into account, the objective to be developed in this work is to understand how the Fourier transform works to be able to recognize a person by means of her voice when having a group of n people.

It is important to first understand the theory of sound and define all the physical properties of the sound in order to have a clearer vision of the operation of voice identification programs.

Subsequently, once the theory of sound has been explained, the theory of the Fourier transform is presented, starting with the definition, its variant in two dimensions and the demonstration of certain theorems that have an important value in the theory of sound.

Then the description of the problem is focused, giving a more detailed description of the Fourier transform. As an important element in this research, a code was generated in Matlab to facilitate the image of a person's voice spectrum.

Finally, the results, conclusions and observations of the data collected in the experimentation of this work are shown.

Keywords: Acoustics, Sound, Spectrum of sound, Convolution, Conjugate.

Índice

Capítulo 1 Introducción.....	1
1.1. Antecedentes	1
1.2. Objetivo general	6
1.2.1. Objetivos específicos	6
1.3. Estructura de la tesis	7
Capítulo II Teoría del Sonido.....	8
2.1. El sonido.....	8
2.2. Propagación y velocidad del sonido	9
2.3. Sonidos periódicos y sus características.....	12
2.4. Espectro de sonido.....	14
2.5. Espectros continuos	16
Capítulo III Teoría de Fourier.....	18
3.1. Series de Fourier.....	18
3.2. Transformada de Fourier unidimensional	21
3.3. Propiedades de la transformada de Fourier	23
Capítulo IV Metodología	28
Capítulo VI Conclusiones	47
Referencias	48

Capítulo 1

Introducción

1.1. Antecedentes

En la actualidad el uso de la voz humana para el control de software y el ingreso de datos en dispositivos electrónicos es una actividad común, gracias a que en los últimos años han existido grandes avances en las diferentes técnicas y herramientas para el desarrollo de esta tecnología (Valenzuela Ramírez, 2015).

El análisis de Fourier es el estudio de las series, de las integrales y de la transformada de Fourier. Se llama así en honor a Joseph Fourier (1768-1830), matemático francés que vivió durante la época napoleónica. Para entender su estudio, se puede decir que las series de Fourier representan funciones periódicas en un intervalo de la recta. Este estudio se puede extender para representar funciones definidas en toda la recta y no periódicas, y en este caso se sustituyen las series por la transformada de Fourier.

Enfocándose en lo que es la transformada de Fourier, se resalta como una herramienta de análisis muy utilizada en el campo científico (acústica, ingeniería biomédica, métodos numéricos, procesamiento de señal, radar electromagnético, comunicaciones, etc.), ya que durante su aplicación transforma una señal de cualquier tipo representada en el dominio del tiempo al dominio de frecuencias sin alterar la información contenida en dicha señal, por lo que se considera una forma diferente de representarla.

Además, a la transformada de Fourier se le puede llamar también espectro de frecuencias, por lo que en acústica, es posible definir un espectro como un gráfico bidimensional de la intensidad del contenido frecuencial de una señal. La definición básica del análisis espectral es que cada sonido complejo puede

descomponerse en un conjunto de ondas sinusoidales simples con diferentes frecuencias y amplitudes. Llamado espectro; estos se representan con la frecuencia en la abscisa (eje horizontal) y la amplitud en la ordenada (eje vertical).

Por otro lado tenemos al sonido, que como todo movimiento ondulatorio es posible representar sus ondas sonoras mediante la transformada de Fourier a manera de una sumatoria de curvas sinusoidales caracterizadas por las mismas magnitudes y unidades de medida que a cualquier onda de frecuencia, como lo es: la amplitud (A), la longitud de onda (λ), la frecuencia (F) y el período (T).

Existe un gran margen de frecuencias entre las cuales se puede generar ondas sonoras, dichas ondas se reducen a los límites de frecuencia que pueden estimular el oído humano para ser percibidas en el cerebro como una sensación acústica. Estos límites de frecuencia se extienden de aproximadamente 20 Hz a cerca 20 KHz y se llaman límites de audición. Las ondas audibles son producidas por cuerdas en vibración o por columnas de aire en vibración.

El reconocimiento de voz, es el proceso por el cual un conjunto de algoritmos computacionales son capaces de traducir fielmente los sonidos de una unidad lingüística (palabra, sílaba o fonema) a un código simbólico que representa al mensaje

Con este mismo principio se elaboró un sistema de reconocimiento de voz a través de un algoritmo, creado en Matlab, que permita la identificación de personas por medio de datos de voz. Para este propósito primero se crea una base de datos que contiene archivos de audio con las voces de n personas dentro de las cuales se encuentra la persona a identificar. Posteriormente se realiza un archivo de audio con la voz de la persona a identificar y se obtiene su espectro; comparando los espectros de la base de datos con el espectro de voz que se desea identificar se puede determinar la identidad de la persona.

Matlab es el acrónimo de Matrix Laboratory, es una herramienta de software que permite realizar cálculos científicos complejos a partir de una representación de

valores basada en matrices. Integran un conjunto de librerías y comandos de alto nivel muy versátil y completo que lo hace idóneo para trabajar con simulaciones dentro de múltiples ámbitos científicos y tecnológicos. Esta herramienta ofrece un entorno interactivo y muy sencillo que permite tanto la programación de algoritmos, la entrada y salida de datos de una forma sencilla y funcional, así como la visualización de éstos. De esta forma posibilita la resolución de problemas de una manera más cómoda mediante el uso de lenguajes de propósito general y tradicional como c y c++ al integrar un poderoso conjunto de funciones y librerías (Óscar Reinoso García, 2018).

Esta herramienta matemática es de uso muy común hoy en día en investigaciones científicas o técnicas, en particular con los estudios relativos a la ingeniería. En este sentido, constituye un software utilizado en multitud de universidades y centros de investigación de todo el mundo y de temáticas muy diversas (Óscar Reinoso García, 2018).

Por lo anterior se considera que la aplicación de la transformada de Fourier es lo más importante en el tratamiento del sonido, ya que sigue con el ejemplo de los programas de reconocimiento de canciones como el ya mencionado Shazam que hace un espectrograma en tiempo real para identificar una canción, realizando comparaciones con cada uno de los espectrogramas de las canciones que tiene almacenados en su base de datos. El objetivo principal es encontrar los picos de frecuencia parecidos a la canción que se está buscando. Utilizando ese mismo principio se creó para esta tesis un código en Matlab que desarrolle ese mismo proceso, pero con la diferencia de identificar a una persona por medio de voz.

A continuación se mencionarán algunas características del sistema auditivo humano.

- **El oído y el proceso de audición:** Desde el punto de vista de la física el oído tiene la función de servir como un transductor de vibraciones que capta

una perturbación del medio, haciendo ciertos procesos de tal manera que propaga, modifica y transforma la señal sonora en señal eléctrica para enviarla al cerebro, el cual la procesa, la interpreta y da sentido. Por lo que para que exista sonido intervienen tres dispositivos: una fuente emisora la cual va a producir una perturbación, un medio por la cual se propagará dicha perturbación y un receptor el cual pueda interpretarla (Rodríguez, 2014). El receptor humano para percibir el sonido es por supuesto el oído que se divide en tres partes: oído externo, oído medio y oído interno.

- **Oído externo** se encarga de captar las ondas sonoras y dirigir las hacia la membrana timpánica, la cual consta de un pabellón auricular, que tiene una estructura con forma de pantalla captadora y tiene además un conducto auditivo externo, éste es una formación tubular que se introduce en el hueso temporal y está cerrada en su extremo interno por la membrana timpánica (Rodríguez, 2006). En la Figura 1 se muestran las partes externas del oído.

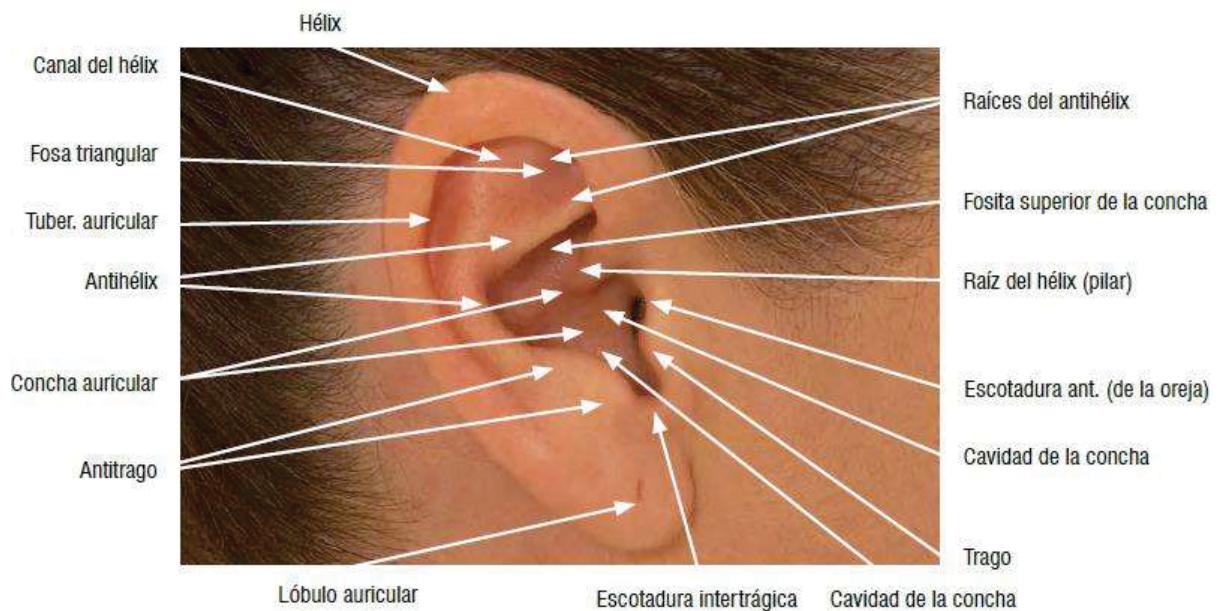


Figura 1: Señalamiento de las partes del oído externo. Imagen tomada del libro Manual de audioprotesismo.

- **Oído medio** es un sistema cavitario, par y simétrico, contenido en su totalidad en el espesor del hueso temporal, excepcionando el segmento faríngeo de la trompa. Por dentro a esta porción media del oído se encuentra el oído interno excavado en el peñasco del temporal. El oído medio está compuesto por la caja timpánica, el sistema neumático temporal y la trompa de Eustaquio (Rodríguez, 2006).
- **Oído interno** Se encuentra ubicado en el interior del peñasco del temporal dentro de la caja del tímpano. Por la complejidad de su forma y estructura se le da el nombre de laberinto, constituido por una serie de espacios excavados en el hueso y en comunicación unos con otros es un laberinto membranoso formado por unas estructuras membranosas alojadas dentro del laberinto óseo (Rodríguez, 2006). En la figura 2 muestra como está conformado el oído medio e interno.

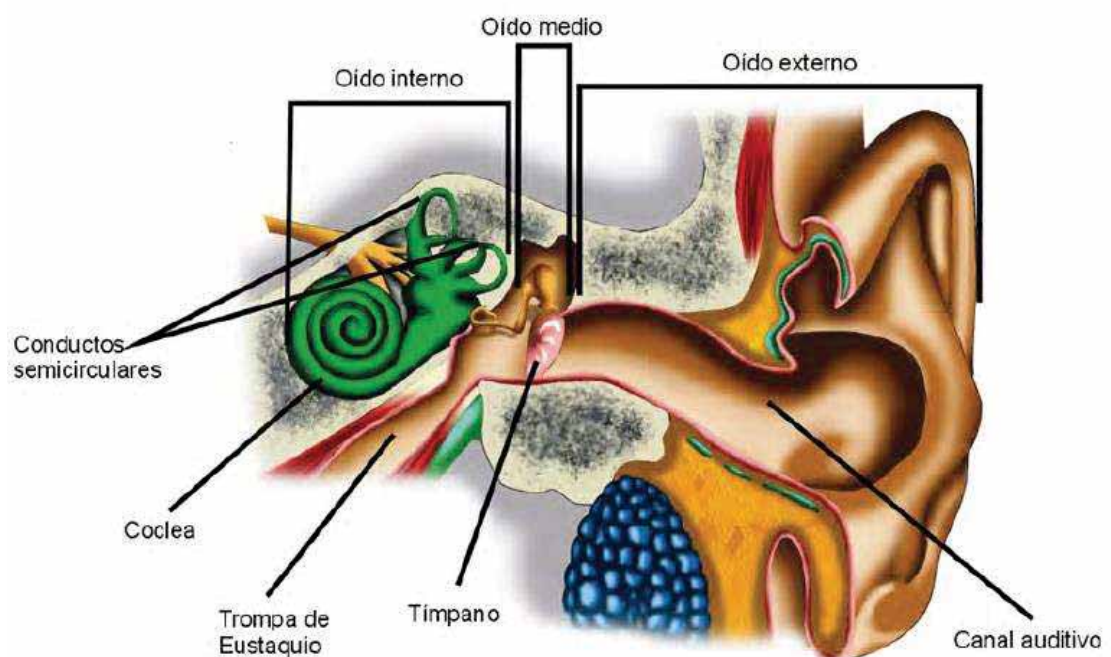


Figura 2: Señalamiento de cómo está conformado el oído medio e interno. Imagen tomada del libro Manual de audioprotésismo.

- **La cóclea** es el órgano periférico de la audición, en ella las señales acústicas se convierten en energía eléctrica o también en señales nerviosas, también en este mismo órgano se discriminan los distintos sonidos según su frecuencia y se codifican los estímulos (Rodríguez, 2006). A continuación se presentan los objetivos de esta tesis.

1.2. Objetivo general

- Identificar a una persona por medio de su voz entre un grupo de personas utilizando la transformada de Fourier e implementando un código en Matlab.

1.2.1. Objetivos específicos

- Analizar los conceptos generales del sonido.
- Comprender la teoría de Fourier.
- Comprender el reconocimiento de voz por medio de la transformada de Fourier en su uso práctico.
- Crear un algoritmo en Matlab que logre identificar a una persona por medio de su voz.

1.3. Estructura de la tesis

En el capítulo dos se plasma información de la teoría del sonido, sus propiedades físicas, la velocidad a la que viaja una onda de sonido y los medios por los que puede viajar. Con estos datos se tendrán elementos para comprender la teoría del sonido y será más fácil entender que es un espectro de sonido.

En el capítulo tres se da una explicación extensa de lo que es la transformada de Fourier, también se presentan las series de Fourier, se hace mención de las fórmulas unidimensionales de la transformada de Fourier, cómo se define un espectro de sonido y qué relación tiene con la transformada de Fourier.

En el último capítulo se dan los resultados obtenidos y se mencionan los problemas al realizar el experimento.

Capítulo II

Teoría del Sonido

2.1. El Sonido

El sonido consiste en la propagación de una perturbación en el aire. Para comprender mejor este concepto se considera un tubo muy largo lleno de aire con un pistón en un extremo. El aire está formado por una cantidad muy grande de pequeñas partículas, manteniendo un equilibrio dinámico, lo cual significa que las partículas no están quietas, sino que se mueven en todas las direcciones debido a la agitación térmica pero con la particularidad de que están homogéneamente repartidas en el interior del tubo. Entonces se desplaza rápidamente el pistón hacia el interior del tubo. Las moléculas que se encuentran junto al pistón serán empujadas por este, mientras que las que se encuentran muy alejadas no. Esto implica que en la zona del pistón el aire se encontrará más comprimido que lejos de él, es decir que la misma cantidad de aire ahora ocupa menos espacio. En otras palabras habrá ahora más partículas por centímetro cúbico cerca del pistón que lejos de él. Por lo que el aire comprimido tiende a descomprimirse desplazándose hacia la derecha, y comprimiendo a su vez el aire que se encuentra próximo a él. Por lo que el proceso se repite así de forma permanente, con lo cual la perturbación original se propaga a lo largo del tubo alejándose de la fuente de la perturbación.

Este proceso se denomina propagación de una onda sonora, siendo similar a lo que sucede cuando en un estanque de agua que está en calma se deja caer una piedra. En el instante en que la piedra golpea el agua, se produce una perturbación que se propaga en forma de una circunferencia cuyo radio va aumentando.

2.2. Propagación y velocidad del sonido

Las ondas de sonido son las que se propagan a través de un material (sólido, líquido, gaseoso). La velocidad de propagación depende de las propiedades elásticas e inerciales del medio. Hay dos tipos básicos de ondas: transversales y longitudinales. En las longitudinales el desplazamiento de las partículas es paralelo a la dirección de propagación, mientras que en las ondas transversales es perpendicular (Miyara, 2003).

Una pregunta muy usual al trabajar con sonido es qué tan rápido se aleja la onda de la fuente de emisión. La respuesta es que el sonido se propaga con una velocidad c que en el aire a una temperatura de $23\text{ }^{\circ}\text{C}$ vale $c=345\text{ m/s}$ o bien $c=1242\text{ km/h}$.

La propagación del sonido está relacionada con variables físicas propias del material como la densidad, la temperatura, la elasticidad, presión, salinidad, etc. En el caso de los medios gaseosos como el aire, las vibraciones son transmitidas de un punto a otro mediante choques entre las partículas que constituyen el gas. De este modo cuando mayor sea la densidad del gas, mayor será la rapidez de la onda.

La rapidez del sonido varía un poco con la temperatura en los sólidos y líquidos, sin embargo en los gases, aumenta con la temperatura por que incrementa la probabilidad de los choques entre moléculas. A nivel molecular con alta elasticidad (rígido) se caracteriza por grandes fuerzas entre sus moléculas. Esto hace que las partículas vuelvan rápidamente a sus posiciones de equilibrio y estén dispuestas a iniciar de nuevo el movimiento, lo que les permite vibrar a altas velocidades, por lo que el sonido viaja más rápido a través de medios con mayor elasticidad.

La velocidad del sonido en sólidos V_s está dada por:

$$V_s = \sqrt{\frac{E}{\rho}}, \quad (1.1)$$

donde E es módulo de Young, ρ es la densidad del material.

El módulo de Young o módulo de elasticidad se define como la fuerza de unión entre los átomos de un metal o aleación. Para los metales, la máxima deformación elástica suele ser inferior a 0.5 por ciento. En general los metales y aleaciones muestran una relación lineal entre la tensión aplicada y la deformación producida en la región elástica del diagrama convencional que se describe por la ley de Hooke.

Dada la siguiente formula se determina el valor del módulo de Young E

$$\sigma = E\varepsilon, \quad (1.2)$$

o

$$E = \frac{\sigma}{\varepsilon}, \quad (1.3)$$

donde σ es la tensión nominal, ε es la deformación nominal. La tensión nominal σ y la deformación nominal ε se define mediante la relación:

$$\sigma = \frac{F}{A_0}, \quad (1.4)$$

y

$$\varepsilon = \frac{l_i - l_0}{l_0} = \frac{\Delta l}{l_0}, \quad (1.5)$$

donde A_0 es el área original de la sección transversal del material antes de iniciarse el ensayo, l_0 es la distancia original entre marcas calibradas y l_i es la distancia entre las mismas después de haberse aplicado una fuerza F .

Mientras que la velocidad del sonido en los líquidos V_l está dada por la siguiente ecuación:

$$V_l = \sqrt{\frac{\beta}{\rho}}, \quad (1.6)$$

donde β es el módulo de compresibilidad de un líquido y ρ es la densidad del material.

El coeficiente de compresibilidad volumétrico se define como la relación del cambio de presión (ΔP) al cambio relativo de densidad (ΔV), mientras que la temperatura se mantiene constante y se da por la siguiente relación,

$$\beta = V \frac{\Delta P}{\Delta V}, \quad (1.7)$$

Por otra parte, la velocidad del sonido en gases V_g se da por la ecuación 8.

$$V_g = \sqrt{\frac{\gamma P}{\rho}}, \quad (1.8)$$

donde γ es el coeficiente de dilatación adiabática, P es la presión del gas y ρ es la densidad del material.

El coeficiente de dilatación adiabática es la razón entre la capacidad calorífica a presión constante (C_p) y la capacidad calorífica a volumen constante (C_v). A veces también conocida como factor de expansión y está dada por la siguiente relación.

$$\gamma = \frac{C_p}{C_v}, \quad (1.9)$$

donde el valor C es la capacidad calorífica o capacidad calorífica específica de un gas. Los sufijos p y v se refieren a las condiciones de presión y de volumen constante respectivamente.

2.3. Sonidos periódicos y sus características

Los fenómenos sonoros analizados en esta tesis consisten en una única perturbación del aire. La mayor parte de los sonidos de la naturaleza son en realidad el resultado de múltiples perturbaciones sucesivas. Estos sonidos se denominan periódicos y pueden dividirse en ciclos, donde cada ciclo abarca todo lo que sucede entre dos perturbaciones sucesivas en el aire.

Siguiendo la analogía de la piedra que cae en un estanque de agua, se puede pensar en una sucesión de rocas que caen sobre la superficie del agua, lo cual dará lugar a una serie de círculos concéntricos que van agrandándose a medida que van surgiendo nuevos círculos. Análogamente, al aire libre y lejos de toda superficie capaz de reflejar el sonido, las sucesivas perturbaciones se propagarán como esferas concéntricas crecientes que se alejan de la fuente que las creó. En presencia de superficies reflectoras, la onda deja de ser esférica para volverse sumamente compleja.

La amplitud de un sonido no es necesariamente constante, sino que puede variar en el tiempo. La mayor parte de los sonidos reales tienen amplitud variable. Se define envolvente de un sonido como la forma que se obtiene uniendo las amplitudes de los ciclos sucesivos.

El tercer parámetro, uno de los más fundamentales en acústica, es la frecuencia f . Se define como la cantidad de ciclos por segundo y se expresa en Hertz (Hz). Esta unidad es equivalente a ciclos por segundo (cps). La frecuencia de los sonidos está comprendida entre los 20 Hz (sonidos graves) y los 20,000 Hz (sonidos agudos) o 20 KHz (Kilohertz). La frecuencia se relaciona con el período T , que se define como el tiempo transcurrido entre una perturbación y la siguiente. Se mide en segundos (s) o milisegundos (ms). El periodo de los sonidos audibles para el ser humano varía entre 0.5 ms (sonidos muy agudos) y 50 ms (sonidos graves).

Existen algunas relaciones matemáticas entre estos parámetros, así el periodo T y la frecuencia f están relacionados por la ecuación

$$f = \frac{1}{T}, \quad (1.10)$$

En las cuales si T se expresa en segundos (s), entonces f se expresa en Hz, si T se expresa en ms, f se expresa en kHz.

Otro parámetro relacionado con los sonidos periódicos, es la longitud de onda, que se representa con la letra griega λ , y es la distancia entre dos perturbaciones sucesivas en el espacio. Se mide en metros (m) o centímetros (cm) y para todos los sonidos audibles está comprendida entre 2 cm (sonidos muy agudos) y 17 m (sonidos muy graves).

La longitud de onda es importante en varias situaciones. En primer lugar, un objeto grande comparado con la longitud de onda es capaz de alterar significativamente la propagación del sonido cuando se interpone entre la fuente sonora y el receptor, así por ejemplo, los sonidos graves pueden “doblar esquinas” fácilmente porque su longitud de onda es grave. Los agudos, en cambio cuya

longitud de onda puede ser apenas algunos cm, se ven considerablemente atenuados.

2.4. Espectro de sonido

Cualquier sonido puede representarse como la suma de una serie de armónicos, es decir que son sonidos senoidales cuyas frecuencias son f , $2f$, $3f$, $4f$, $5f$, etc. Por ejemplo, la nota musical “La” del piano, cuya frecuencia es de 440 Hz, contiene armónicos de frecuencias 440 Hz, 880 Hz, 1320 Hz, 1760 Hz, 2200 Hz, etc. Cada uno de estos armónicos puede tener su propia amplitud. En la Figura 3(a) se muestran los primeros armónicos de una onda cuadrada, y en la Figura 3(b) se ha obtenido su suma, que según se aprecia se va aproximando a la onda cuadrada.

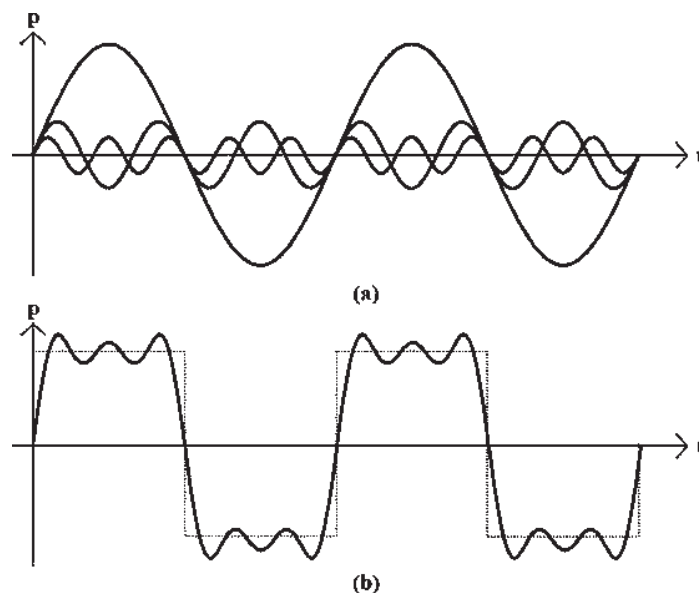


Figura 3: (a) Los tres primeros armónicos no nulos de una onda cuadrada de frecuencia f_0 , cuyas frecuencias son f_0 , $3f_0$ y $5f_0$. (b) El resultado de superponer los tres armónicos, comparado con la onda cuadrada. Si bien tres armónicos son poca cantidad, se observa que comienza a esbozarse la forma de la onda cuadrada. Imagen tomada del libro de acústica y sistemas de sonido.

La información sobre las frecuencias que contienen un determinado sonido y sus respectivas amplitudes constituyen lo que se denomina el espectro del sonido. El espectro se puede especificar en forma de tabla o se puede representar gráficamente mediante un espectrograma que es un gráfico con dos ejes: el horizontal, graduado en frecuencias y el vertical en amplitud.

En la Tabla 1 se indican los primeros armónicos para las ondas cuadradas, triangular y diente de sierra, suponiendo que la amplitud es 1 en los tres casos.

En la Figura 4 se ha representado el espectro de sonido de una onda cuadrada de amplitud 1 y de frecuencia 100 Hz incluyendo hasta el armónico.

Tabla 1 Amplitud de los primeros 7 armónicos del espectro de las ondas cuadrada, triangular y diente de sierra. Tabla tomada del libro de acústica y sistemas de sonido.

No. De armónico	Cuadrada	Triangular	Diente de Sierra
1	1.27	0.81	0.64
2	0	0	0.32
3	0.42	0.09	0.21
4	0	0	0.16
5	0.25	0.032	0.13
6	0	0	0.11
7	0.18	0.017	0.091

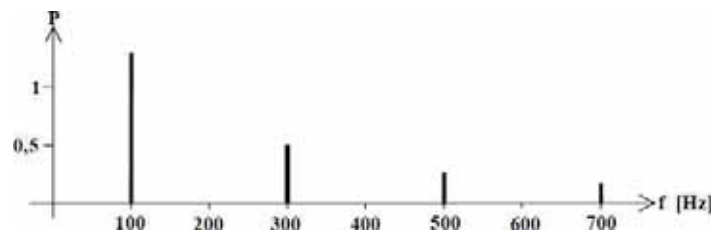


Figura 4: Espectro de una onda cuadrada de amplitud 1 y de frecuencia 100 Hz. Esta onda tiene únicamente armónicos impares. Imagen tomada del libro de acústica y sistemas de sonido.

Así como la amplitud de un sonido puede variar en el tiempo de acuerdo con su envolvente, también es posible que los diversos armónicos que integran determinada forma de onda posean sus correspondientes envolventes, que no tienen por qué ser iguales.

2.5. Espectros continuos

Existe aún otro tipo de sonidos, formados por una cantidad muy grande de frecuencias que están muy próximos entre sí y que se denominan genéricamente ruido. Algunos ejemplos de esto son el sonido del mar y el sonido que se emite al pronunciar las consonantes f, j, s, z o simplemente al soplar. Debido a la gran cantidad de frecuencias y al hecho de que cada uno es de amplitud muy pequeña, lo más conveniente es representar como una curva continua como se muestra en la figura 5 y se denomina como densidad espectral.

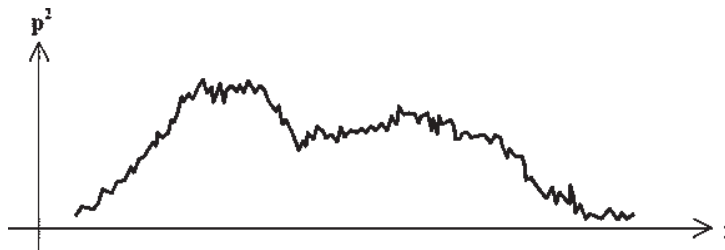


Figura 5: Ejemplo de espectro continuo de un ruido. En el eje horizontal se indica la frecuencia y en el vertical la densidad espectral, que representa la energía en función de la frecuencia. Imagen tomada del libro de acústica y sistemas de sonido.

Existen dos tipos de ruido que tienen importancia específica en Acústica: el ruido blanco y el ruido rosa. El ruido blanco (Figura 6a) se caracteriza por tener una densidad espectral constante es decir es igual para todas las frecuencias. Esto significa que contiene todas las frecuencias con igual amplitud. El nombre de ruido “blanco” proviene de realizar una analogía de la luz blanca, que contiene a todos los colores del espectro con la misma intensidad. El ruido rosa (Figura 6b) contiene mayor proporción de bajas frecuencias. Tiene la particularidad de que en cada octava tiene la misma energía sonora. El ruido rosa tiene aplicación en la ecualización de sistemas de sonido mediante ecualizadores por octavas o por tercios de octava. Es también una señal útil para la prueba de equipos de sonido, ya que es un tipo de ruido que suena natural al oído.

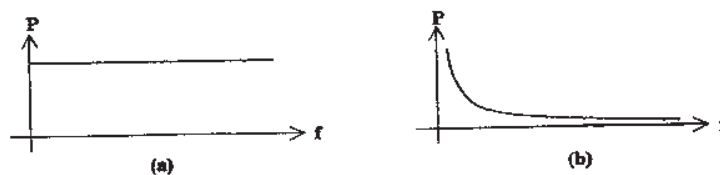


Figura 6: (a) Densidad espectral del ruido blanco. (b) Densidad espectral del ruido rosa. Imagen tomada del libro de acústica y sistemas de sonido.

Capítulo III

Teoría de Fourier

En la física, dentro del análisis de fenómenos, se trabaja con sistemas referentes al análisis de series temporales, tratamiento digital de imágenes, tratamiento de señales de audio o sistemas que oscilan o vibran. Una de las herramientas matemáticas más importantes en estas aplicaciones es la transformada de Fourier ya que juega un papel importante en el análisis y modelado de tales problemas. El análisis de Fourier representa un esquema sistemático con este propósito. Si una función periódica $f(t)$ de periodo T se puede representar por:

$$f(t) = f(t + T), \quad (2.11)$$

donde la constante T es el periodo y si esta cumple con las condiciones de Dirichlet (Goodman, 2005), entonces se puede aplicar la teoría de Fourier.

3.1. Series de Fourier

Una función $g(x)$ cualquiera con periodo T , se puede representar mediante la serie trigonométrica de Fourier:

$$g(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi nx}{T}\right) + b_n \sin\left(\frac{2\pi nx}{T}\right) \right], \quad (2.12)$$

o bien se puede expresar mediante $f_x = \frac{1}{T}$ que es la frecuencia fundamental, por lo que se obtiene:

$$g(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \text{Cos}(2\pi n x f_x) + b_n \text{Sen}(2\pi n x f_x)], \quad (2.13)$$

donde los coeficientes a_n y b_n son las amplitudes de cada una de las componentes y estas determinan la importancia o factores de peso de las diversas contribuciones de frecuencia espacial angular, esto es a_n y b_n son las transformada de Fourier del Coseno y del Seno dadas por:

$$a_n = \frac{1}{T} \int_{-T}^T g(x) \text{Cos}(2\pi x f_x) dx, \quad (2.14)$$

$$b_n = \frac{1}{T} \int_{-T}^T g(x) \text{Sen}(2\pi x f_x) dx, \quad (2.15)$$

si la función $g(x)$ es real, los coeficientes son también reales por lo que estos pueden ser calculados después de integrar sobre un periodo completo la expresión analítica.

Se puede observar que los componentes de la frecuencia tienen una separación constante igual a la frecuencia fundamental f_x . Si la función es simétrica ($g(x) = g(-x)$), entonces solo los coeficientes a_n pueden ser diferentes de cero, por otra parte si es anti simétrica ($g(x) = -g(-x)$) entonces serán los coeficientes b_n pueden ser diferentes de cero. En la Figura 7 se muestran algunas de las funciones periódicas más comunes.

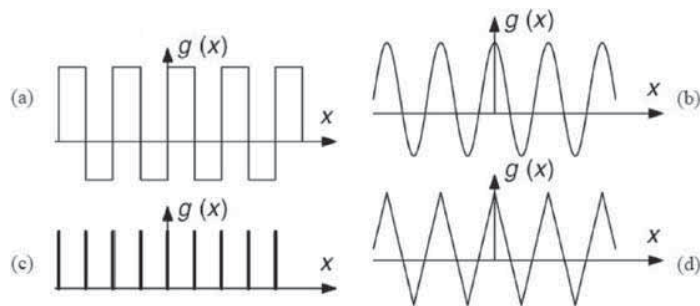


Figura 7: Funciones periódicas.

Las series de Fourier también pueden ser escritas en términos de funciones complejas donde los coeficientes a_n y b_n deben ser complejos. Usando las ecuaciones de Euler se obtiene:

$$\text{Cos}(2\pi x f_x) = \frac{e^{2\pi i n x f_x} + e^{-2\pi i n x f_x}}{2}, \quad (2.16)$$

$$\text{Sen}(2\pi x f_x) = \frac{e^{2\pi i n x f_x} - e^{-2\pi i n x f_x}}{2i}, \quad (2.17)$$

por los que la ecuación (2.13) puede ser escrita de la siguiente manera:

$$g(x) = \sum_{n=-\infty}^{\infty} C_n e^{2\pi i n x f_x}, \quad (2.18)$$

donde los coeficientes C_n pueden ser reales, imaginarios o complejos. Estas funciones exponenciales son también ortogonales ya que son funciones trigonométricas. Entonces los coeficientes pueden ser calculados como:

$$C_n = \int_{-T}^T g(x) \cos(2\pi x f_x) dx. \quad (2.19)$$

En este caso, los coeficientes C_n corresponden a los valores positivos de la amplitud. En general estos coeficientes son complejos.

3.2. Transformada de Fourier unidimensional

Si el periodo de $g(x)$ se incrementa y la separación de las componentes decrece, en el límite cuando el periodo tiende a infinito, el intervalo de frecuencia entre armónicos tiende a cero. Cualquier función continua no periódica puede ser considerada como una función periódica con un periodo infinito y por lo tanto, se puede representar como una serie de Fourier, es decir, se puede representar por una suma infinita de funciones sinusoidales. Por consecuencia, la serie de la ecuación (2.13) se puede representar como una integral y con esto llegamos al concepto de Transformada de Fourier para cualquier función continua no periódica.

Sea $g(x)$ una función integrable definida en \mathbb{R} . Su transformada de Fourier o integral de Fourier unidimensional será la función definida también en \mathbb{R} , que se representa como $G(f)$ dada por:

$$G(f) = \int_{-\infty}^{\infty} g(x)e^{-2\pi if_x} dx. \quad (2.20)$$

La condición para que exista $G(f)$ generalmente está dada por:

$$\int_{-\infty}^{\infty} |g(x)dx| < \infty. \quad (2.21)$$

La integral del valor absoluto de $g(x)$ debe ser finita.

La transformada de Fourier $G(f)$ es el espectro de amplitud de la función $g(x)$ o espectro de Fourier y su amplitud es el módulo al cuadrado de este espectro. Un ejemplo de una transformada de Fourier se tiene al utilizar una función sinusoidal $g(x)$ la cual presenta un espectro de Fourier único con dos frecuencias características que se pueden interpretar como su huella digital. Este espectro es un par de funciones delta de Dirac localizadas simétricamente con respecto al origen del plano de frecuencia correspondiente. Dada $G(f)$, la función $g(x)$, puede ser obtenida a través de la transformada de Fourier, definida por:

$$g(x) = \int_{-\infty}^{\infty} G(f)e^{2\pi if_x} df. \quad (2.22)$$

Aquí x es la variable espacial y su dominio se refiere como dominio espacial. Por otro lado, f es una variable de frecuencia, por lo que su dominio es el de frecuencias espaciales o dominio de Fourier. Un par de Transformadas de Fourier

está definido por las ecuaciones (2.20) y (2.22). Ambas funciones $g(x)$ y $G(f)$, pueden ser reales o complejas.

3.3. Propiedades de la transformada de Fourier

La definición básica de la transformada de Fourier conduce a una rica estructura matemática asociada con dicha operación. Las propiedades que se presentan a continuación como teoremas matemáticos son extensiones directas de los enunciados unidimensionales. Por lo que a continuación se hará mención de las propiedades de la transformada de Fourier, de las cuales solo se demostrará el primer teorema de forma gráfica debido a la importancia en el tratamiento del sonido. Los demás teoremas solo se hará mención y no se demostrarán debido a que en el tratamiento de sonido no son muy utilizados o su demostración se sale de los propósitos de este trabajo.

Teorema de linealidad. La transformada de una suma ponderada de dos (o más) funciones es idéntica a la suma ponderada de las transformadas individuales

$$F(\alpha g + \beta h) = \alpha F(g) + \beta F(h). \quad (2.24)$$

Para demostrar este teorema se utilizaron las notas musicales “Do” y “Re” dada por la figura (8) y tomando esta función como la función de entrada en la definición de la transformada de Fourier se obtiene la gráfica mostrada en la figura (9). Usando el programa de Matlab, se implementó un código para crear una escala

musical (una octava en la escala musical) por lo que con dichas notas se determinó su transformada de Fourier, dando como resultado dicha imagen.

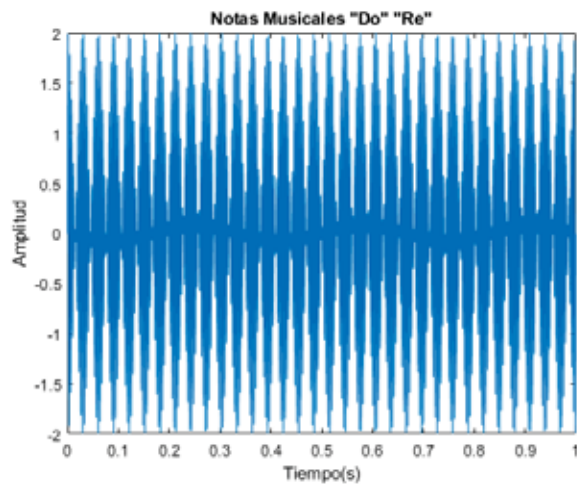


Figura 8: Gráfica de las notas "Do" y "Re".

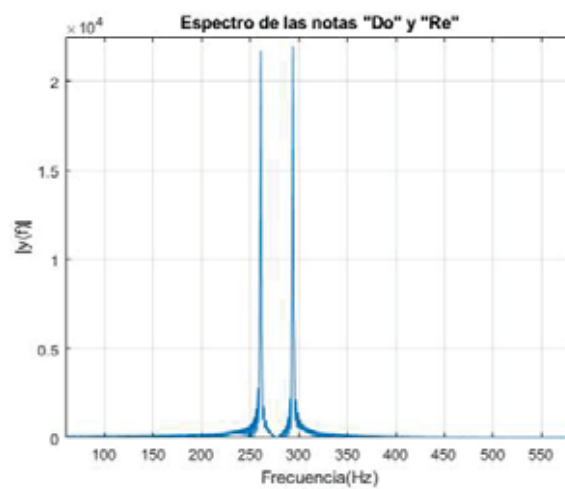


Figura 9: Gráfica de su transformada de Fourier.

Teorema de similaridad. Si $a \in \mathbb{R}$ y $F(g) = G(f_x)$, entonces

$$F[g(ax)] = \frac{1}{|a|} G\left(\frac{f_x}{a}\right). \quad (2.25)$$

La función $g(ax)$ representa una contracción de las coordenadas del dominio de frecuencias f_x por un factor a , además de un cambio en la amplitud del espectro.

Teorema de desplazamiento. Sea $F[g(x)] = G(f_x)$ y sea $a \in \mathbb{R}$, entonces

$$F[g(x - a)] = G(f_x)e^{-i2\pi(f_x a)}. \quad (2.26)$$

Teorema de Rayleigh (Teorema de Parseval). Si $F[g(x)] = G(f_x)$, entonces

$$\int_{-\infty}^{\infty} |g(x)|^2 dx = \int_{-\infty}^{\infty} |G(f_x)|^2 df_x. \quad (2.27)$$

La integral del lado izquierdo se puede interpretar como la energía contenida en la forma de onda $g(x)$. Esto a su vez conduce a la idea de que la cantidad de energía $|G(f_x)|^2$ se puede interpretar como una densidad de energía en el dominio de frecuencia.

Teorema de convolución. Si $F[g(x)] = G(f_x)$ y $F[h(x)] = H(f_x)$, entonces

$$F \left\{ \int_{-\infty}^{\infty} g(\xi)h(x - \xi) \right\} = G(f_x)H(f_x). \quad (2.28)$$

La convolución de dos funciones en el dominio espacial es equivalente a la más simple operación de multiplicar sus transformadas individuales.

Teorema de autocorrelación. Si $F[g(x)] = G(f_x)$, entonces

$$F \left\{ \int_{-\infty}^{\infty} G(\xi)G^*(x - \xi)d\xi \right\} = |G(f_x)|^2, \quad (2.29)$$

Similarmente

$$F|g(x)|^2 = \int_{-\infty}^{\infty} G(\xi)G^*(x - \xi)d\xi. \quad (2.30)$$

Esta teoría puede considerarse como un caso especial del teorema de convolución en el que se hizo la convolución $g(x)$ con $g^*(x)$.

Teorema de la integral de Fourier. En cada punto de continuidad en g ,

$$FF^{-1}g(x) = F^{-1}Fg(x) = g(x). \quad (2.31)$$

En cada punto de discontinuidad de g , las dos transformadas sucesivas producen un promedio angular de los valores de g en una pequeña vecindad alrededor de ese punto. Por lo que la transformada sucesiva y la transformada inversa de una función producen la función original, excepto en los puntos de discontinuidad.

Capítulo IV

Metodología

La transformada de Fourier es de gran importancia para la realización de este trabajo por lo que se implementó un código en el software Matlab para el reconocimiento de voz utilizando esta transformada.

Para realizar el reconocimiento de voz por medio de la transformada de Fourier, se utilizan al menos dos archivos de audio, uno llamado “master” que contiene la voz que se desea identificar y otro que contiene la voz con la que se compara el master; sin embargo, se pueden utilizar más de dos archivos para comparar si la persona que se desea identificar está dentro de una base de datos o no.

Utilizando el archivo master y recordando que la transformada de Fourier calcula el espectro en frecuencias de un archivo de audio, se obtiene el complejo conjugado de este espectro y se utiliza como filtro por el cual pasa la transformada de Fourier del archivo que se va a comparar. Si ambos archivos coinciden (el master y el que se compara) entonces ambos espectros producirán una constante cuya transformada inversa es una función delta de Dirac o también conocida como función impulso. Si los archivos no son iguales (el master y el que se compara) entonces sus transformadas de Fourier son distintas y al pasar una sobre la otra producirán un producto diferente a una constante cuya transformada de Fourier inversa se desvía de la función delta de Dirac. Finalmente, el criterio para distinguir si los dos archivos son iguales o no, se reduce a identificar si se produce un impulso o no.

Sean los archivos de audio master y de comparación representados por las funciones $m(t)$ y $c(t)$ respectivamente cuyas transformadas de Fourier son representados por $M(\xi)$ y $C(\xi)$.

La función que se utilizará como filtro, $F(\xi)$, es el complejo conjugado del espectro del archivo master dado por

$$F(\xi) = M^*(\xi). \quad (3.32)$$

El producto $v(\xi)$ se define entonces como

$$v(\xi) = F(\xi)E(\xi), \quad (3.33)$$

donde $E(\xi)$ es el espectro del archivo de audio con el que se está comparando el audio master.

Si ambos archivos coinciden, entonces $E(\xi) = M(\xi)$ y si no son iguales entonces $E(\xi) = C(\xi)$. Finalmente, se calcula la transformada de Fourier inversa del producto $v(\xi)$ dada por

$$p(x) = \int_{-\infty}^{\infty} v(\xi) e^{i2\pi(\xi x)} d\xi, \quad (3.34)$$

y se calcula su intensidad o irradiancia dada por

$$I(x) = |p(x)|^2. \quad (3.35)$$

Los dos casos posibles que se pueden presentar es cuando el master y el archivo a comparar son iguales, es decir $c(t) = m(t)$, por lo cual el espectro $E(\xi) = M(\xi)$ y el producto dado por la ecuación 3.33 es

$$v(\xi) = M^*(\xi)M(\xi) = \text{constante} \quad (3.36)$$

y entonces, la transformada de Fourier inversa es

$$p(x) = \int_{-\infty}^{\infty} v(\xi)e^{i2\pi(\xi x)}d\xi = \text{delta de dirac}(x), \quad (3.37)$$

finalmente, la irradiancia está dada por

$$I = |\text{delta}|^2 = \text{delta}(x). \quad (3.38)$$

El otro caso que se puede presentar es cuando $c(t) \neq m(t)$. Sea el espectro $E(\xi) = C(\xi)$. El producto $v(\xi)$ está dado por

$$v(\xi) = M^*(\xi)C(\xi), \quad (3.39)$$

cuya transformada de Fourier inversa es

$$p(x) = \int_{-\infty}^{\infty} v(\xi)e^{i2\pi(\xi x)}d\xi = V(x) \quad (3.40)$$

y la irradiancia es

$$I(x) = |V(x)|^2. \quad (3.41)$$

Por lo que se desvía de la función impulso o delta de Dirac indicando que ambas funciones no son iguales.

Por medio del software de Matlab se brinda el soporte para desarrollar un código que implementa la transformada de Fourier, en el que a continuación se muestran los procesos que realizará dicho código.

- Las palabras que se utilizaron fueron las siguientes: Arroz, Azul, Casa, Perro, Carro y Mesa. Esto debido a que las vocales a, e, y o suenan con mayor fuerza por lo que facilitan el reconocimiento de las palabras.
- Se almacenaron las voces de 18 personas, cada persona hizo mención de cada una de las palabras ya antes mencionadas.
- Después se cargó esa información en un código de Matlab con un método de selección, en donde primero muestra una lista de las palabras usadas, luego despliega los nombres de las personas que hicieron mención de dicha palabra.
- Se escoge una palabra y después a una persona y se toma su archivo de voz el cual será nombrado como audio master o principal.
- Los audios son guardados en vectores de tamaño $1 \times n$ manteniendo el mismo tamaño, cada audio tiene dos segundos de duración.
- Se prosigue a seleccionar a una segunda persona y su archivo de audio será nombrado como audio a comparar.
- Se determina la transformada de Fourier de los dos archivos seleccionados.
- Al determinar la transformada de Fourier del primer archivo de voz, se sacó el conjugado de dicha transformada el cual se utilizará como filtro, para después multiplicarlo por el audio a comparar.
- Con dicha multiplicación se obtiene la transformada inversa de Fourier, una vez calculada la transformada inversa se determina la intensidad y así llegar al resultado final el cual nos debe mostrar una función delta o pico. Si lo que muestra no es una función delta es que los archivos de voz son de diferentes personas.

En la Figura 10 se muestra un diagrama de flujo que sintetiza el procedimiento realizado para la identificación de una persona por medio de su voz.



Figura 10: Diagrama de flujo de los pasos que realiza el algoritmo creado.

Para el procesamiento de las grabaciones y el procesamiento de la información se utilizó una computadora Laptop Lenovo x220 Tablet con su micrófono integrado, sistema operativo Windows 10, con el software de Matlab R2019a.

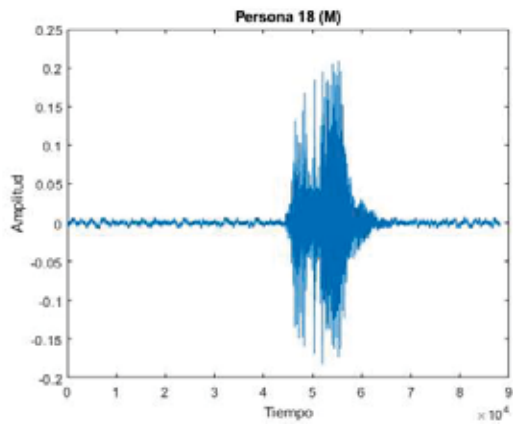
Capítulo V

Resultados

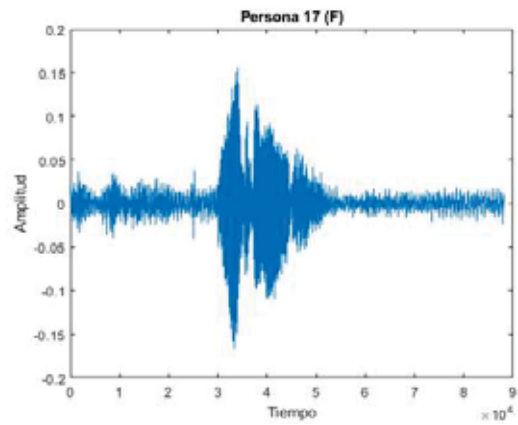
En este apartado se dan a conocer cuáles fueron las problemáticas y resultados que se obtuvieron al momento de poner a prueba el código implementado en Matlab.

Durante las grabaciones de las voces se dieron las siguientes indicaciones: Decir las palabras de una forma normal evitando dar mayor énfasis a otras vocales o alargar las palabras, otra indicación que se dió fue que se dijera la palabra como si mantuviera una conversación con otra persona, ya que la mayoría de personas tendían a gritar o engrosar la voz. También se dijo que se mantuviera un control emocional, debido a que si se presenta alguna emoción durante la grabación esto puede cambiar los resultados al momento de hacer las comparaciones de la transformada de Fourier, ya que el código no es capaz de detectar emociones al momento de decir las palabras.

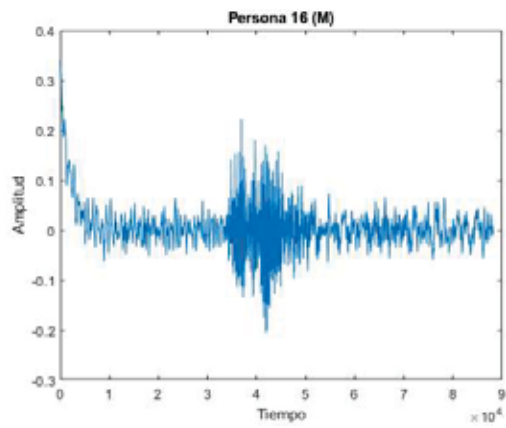
A continuación, se muestran las gráficas de voz de cada persona que participó durante la grabación de los archivos de voz.



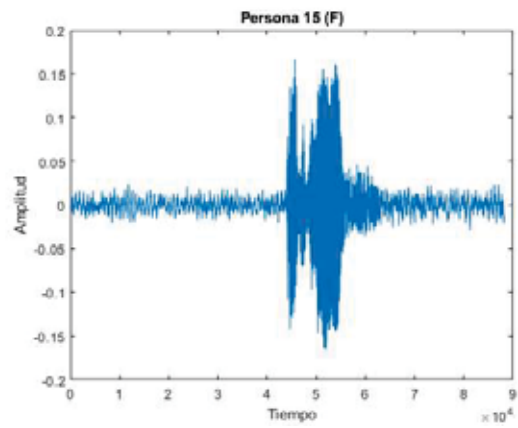
(a)



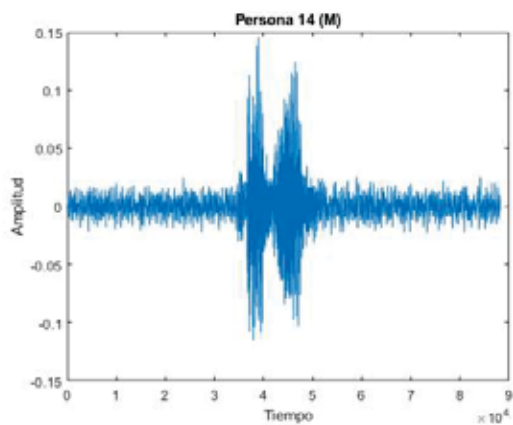
(b)



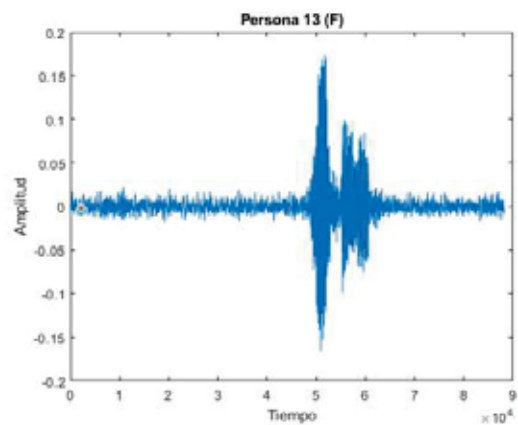
(c)



(d)

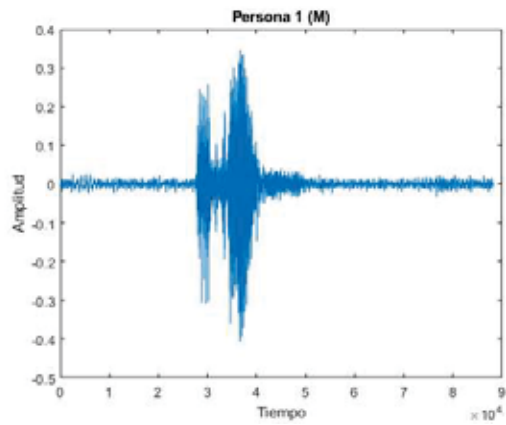


(e)

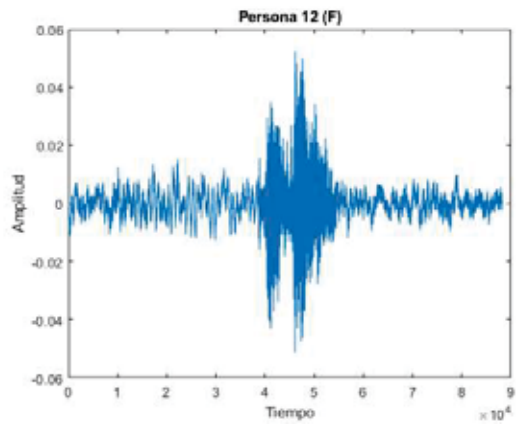


(f)

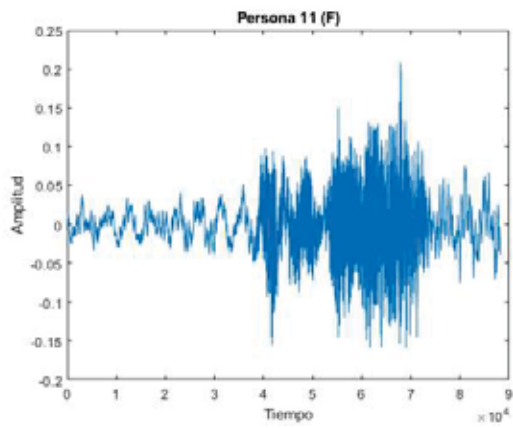
Figura 11: Gráfica de los archivos de voz almacenados



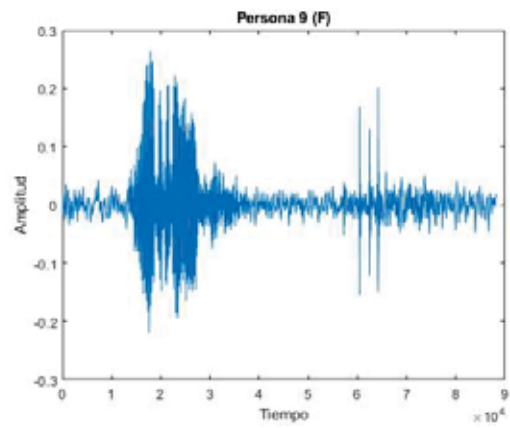
(a)



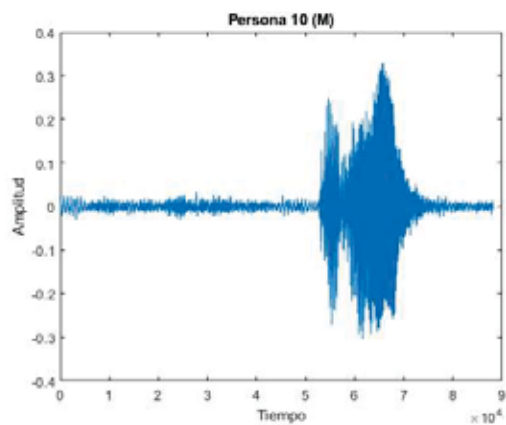
(b)



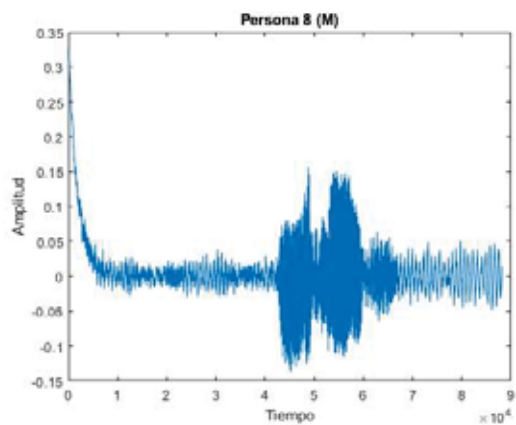
(c)



(d)

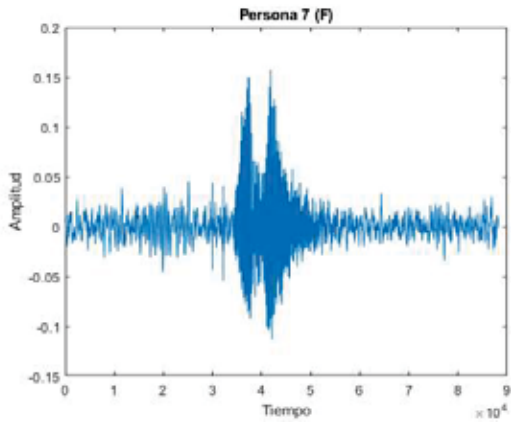


(e)

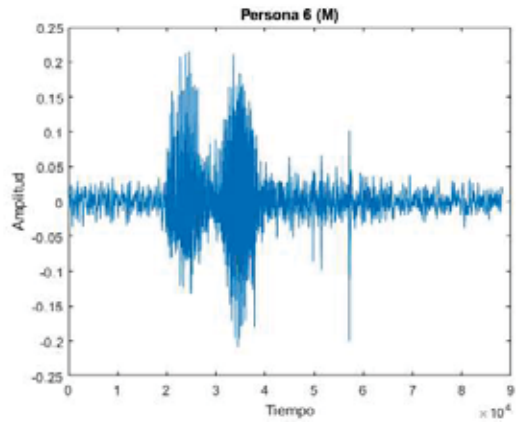


(f)

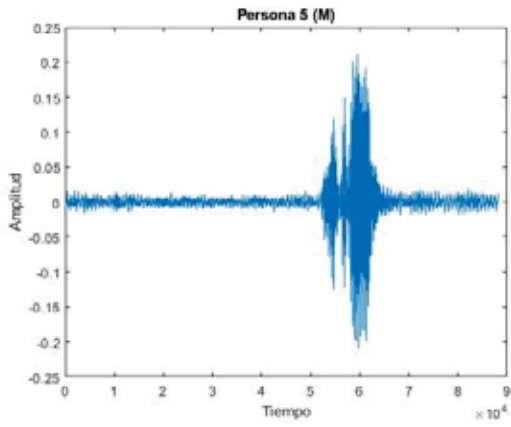
Figura 12: Gráfica de los archivos de voz almacenados



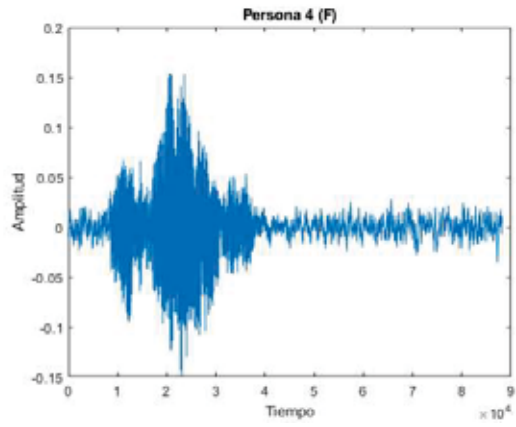
(a)



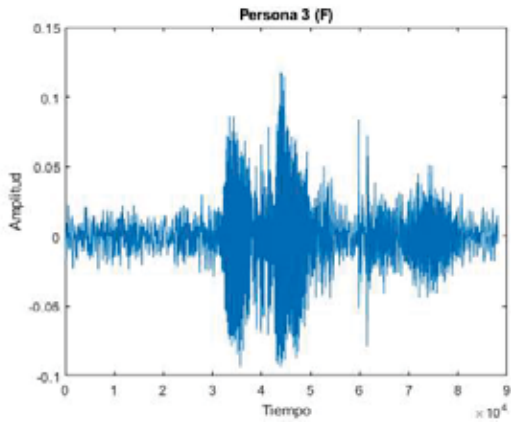
(b)



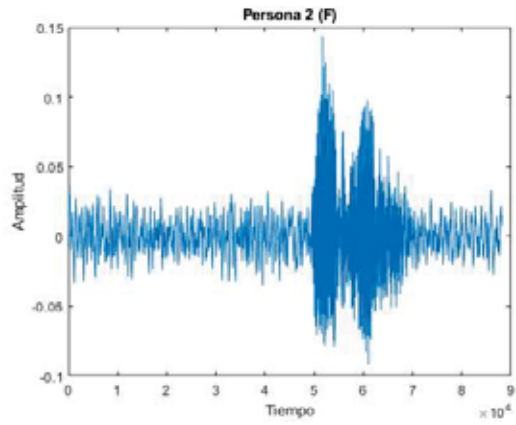
(c)



(d)



(e)



(f)

Figura 13: Gráfica de los archivos de voz almacenados

Durante las grabaciones de cada archivo se trató de grabar en lugares donde no hubiera ruidos externos a los que estaban en la habitación, debido a que estos podían causar interferencia en la realización del reconocimiento de voz.

Siguiendo con la idea principal del código, se puede resumir como la creación de un filtro hecho con la primer transformada de Fourier que corresponde a la voz master.

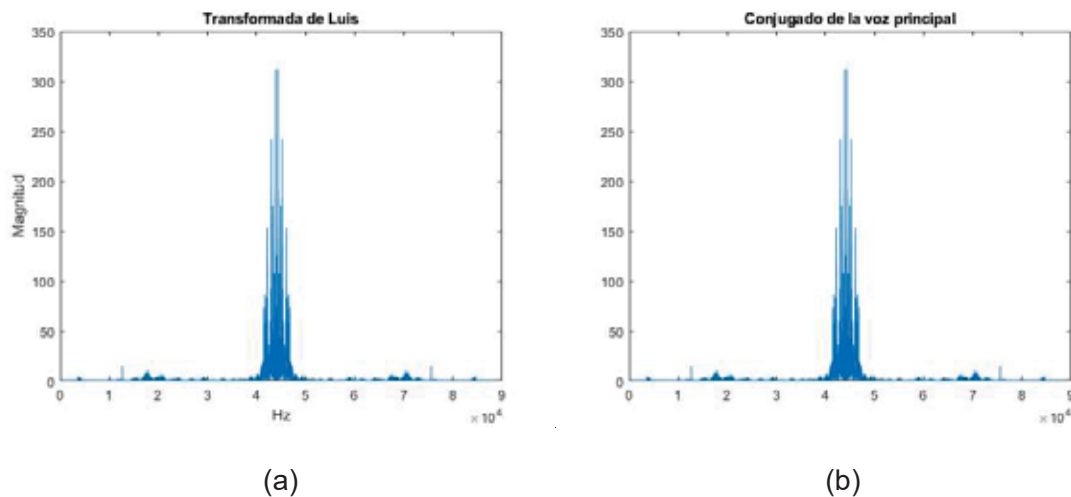


Figura 14: En la figura (a) se muestra la transformada de Fourier del archivo de audio de la voz master y en (b) se muestra el complejo conjugado de la transformada de Fourier

El conjugado de la transformada de Fourier mostrado en la figura 14(b) es el archivo que se usará como filtro, por lo que el siguiente paso es calcular la transformada de Fourier de los archivos de audio con el que se va a comparar con la figura 14(b). A continuación en la figura 15 se muestran dos casos posibles, uno es cuando el archivo de audio a comparar es diferente al del master (figura 15(a)) y por otro lado el audio a comparar coincide con audio el master (figura 15(b)), en la

que se muestra el producto del filtro por la transformada de Fourier del audio al que se compara.

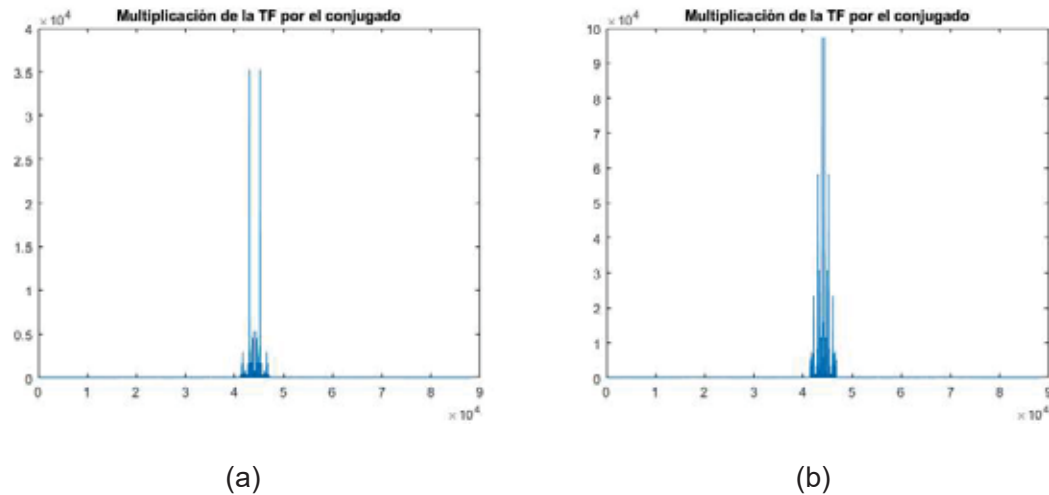
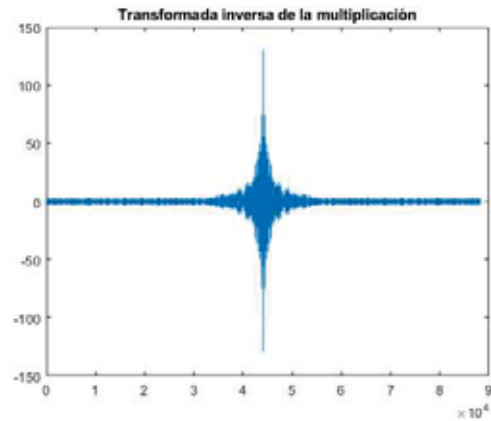


Figura 15: Graficas de la multiplicación del conjugado por la transformada de Fourier del audio a comparar, (a) corresponde cuando los dos archivos de voz son diferentes, (b) es cuando los archivos de audio son el mismo

El siguiente paso es determinar la transformada inversa de Fourier del producto del filtro por la transformada de Fourier de los archivos de audio a comparar, en la figura 16(a) se muestra la transformada inversa cuando el archivo de audio no coincide con el master y en la figura 16(b) se muestra la transformada inversa cuando los archivos de voz coinciden con el master.



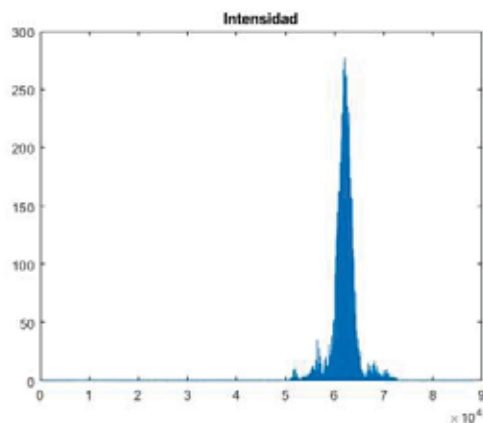
(a)



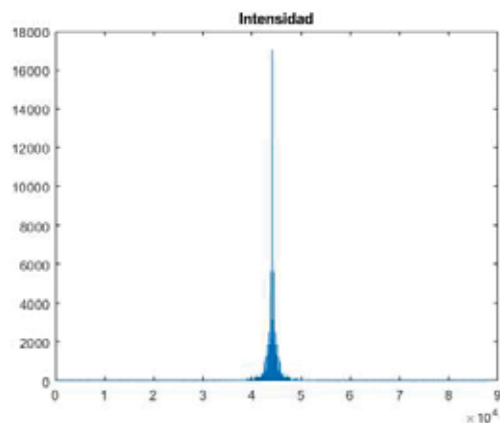
(b)

Figura 16: Representación de la transformada de Fourier inversa de los archivos de voz, (a) los archivos de voz seleccionados son diferentes, (b) los archivos de voz seleccionados son los mismos.

Por último, se determina la intensidad de la transformada Inversa para dar una mejor visualización de la función delta que se está buscando, si ambas voces son las mismas la función delta será visible a simple vista y por otro lado si no lo son la función delta será difícil de observar. En la figura 17 se aprecia este resultado.



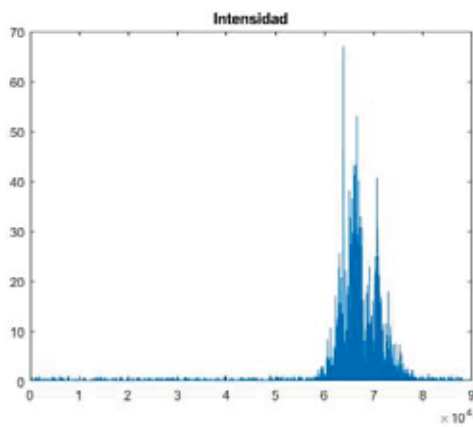
(a)



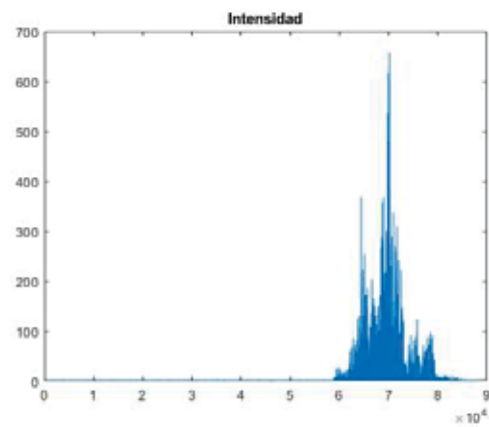
(b)

Figura 17: Se muestran los resultados cuando los archivos de voz son diferentes y cuando son el mismo. (a) Se puede apreciar como es que el pico buscado no se puede visualizar. (b) Se aprecia como la delta que se esta buscando se visualiza a simple vista.

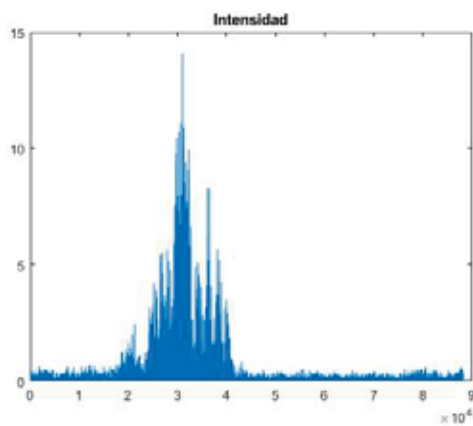
El procedimiento mencionado anteriormente se realizó para cada uno de los archivos de voz, resultando que en ninguno de los casos se vuelve a producir la función delta de Dirac. En la figura 18 se muestran algunos resultados.



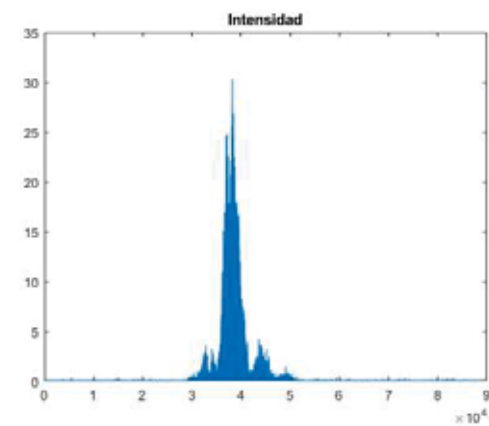
(a)



(b)



(c)



(d)

Figura 18: Resultados del experimento, en donde se compararon voces diferentes y no es posible visualizar la función delta. Se usaron las voces de los usuarios a)1, b)2, c)3, d)4

El código que se realizó en esta tesis fue en Matlab versión 2019a y es manipulado directamente por el usuario, como un trabajo futuro se pretende realizar una interfaz gráfica para que cualquier usuario pueda hacer uso de este software.

Dicho experimento se repitió nueve veces para observar si la grabación podía ser identificada en otro momento. En la figura 19 y en la figura 20 se muestran los resultados obtenidos.

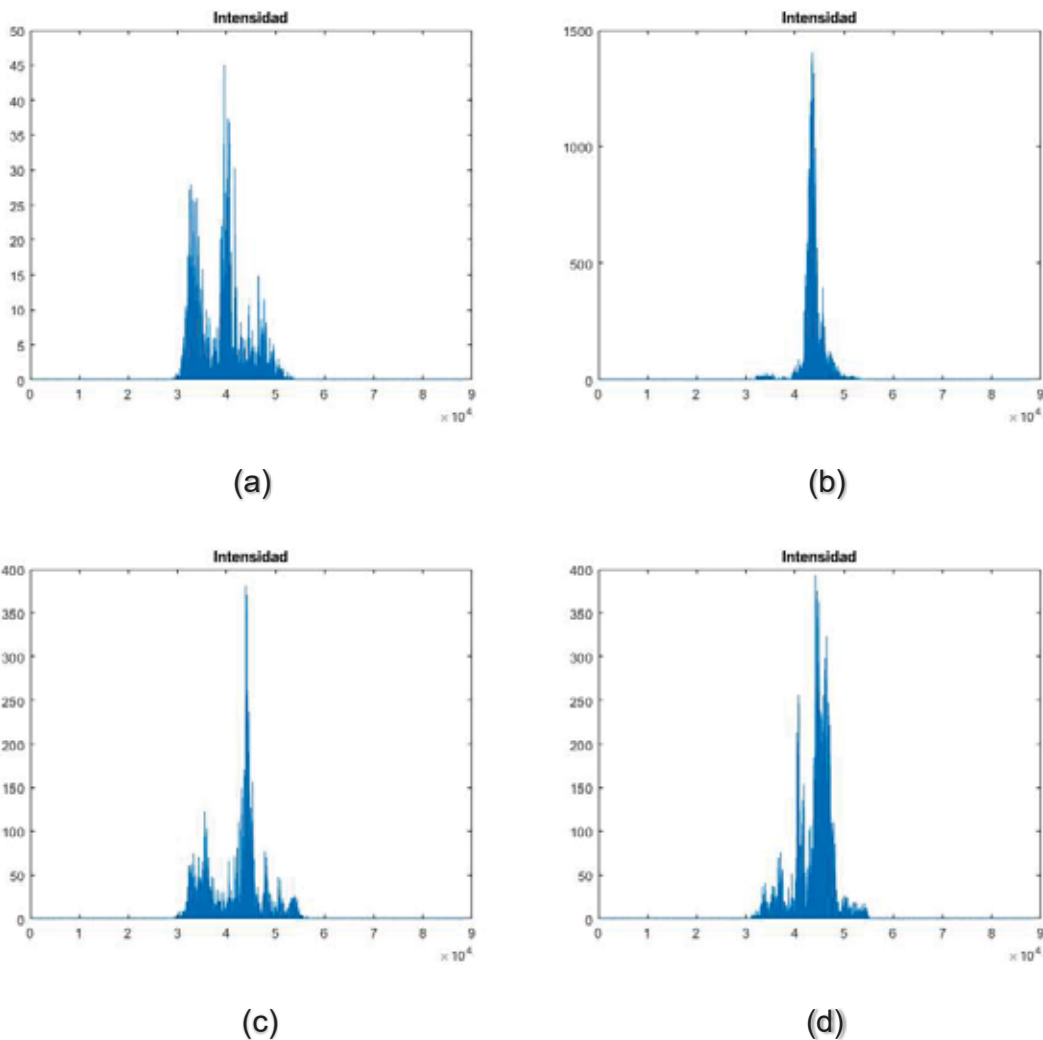
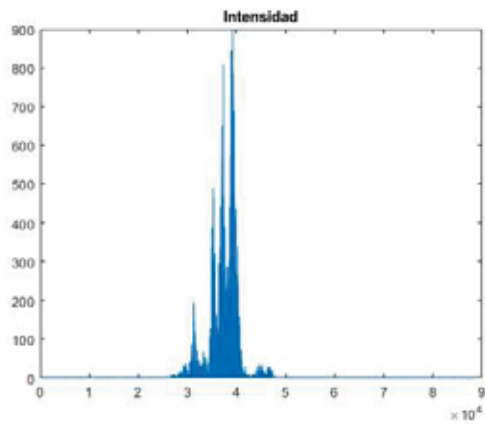
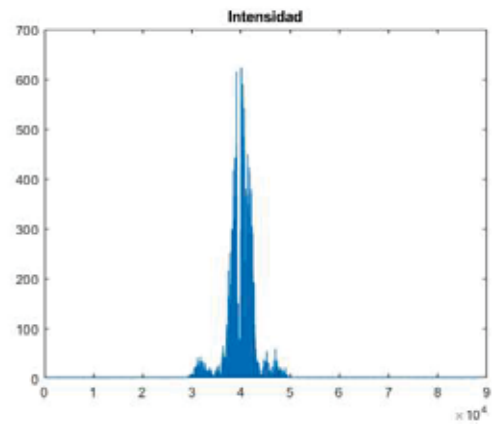


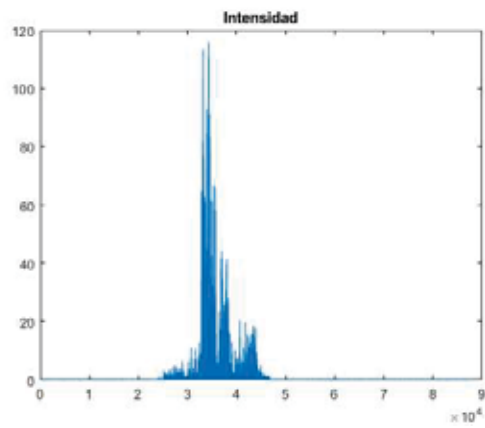
Figura19: Se puede observar que ninguna figura muestra la función delta.



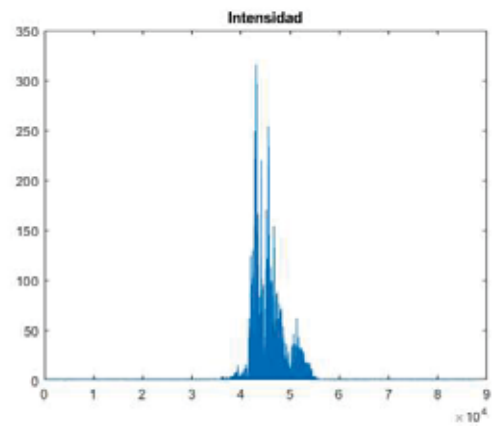
(a)



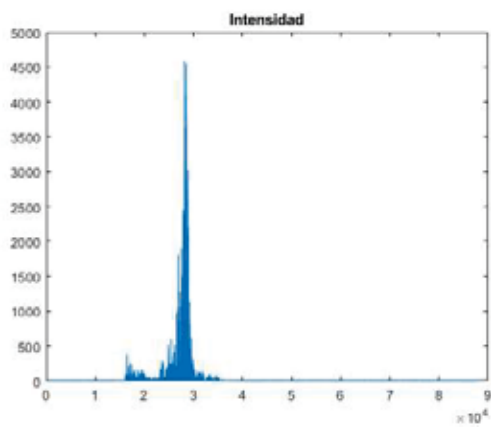
(b)



(c)



(d)



(d)

Figura 20: Se puede observar que ninguna figura muestra la función delta.

Por lo que se puede observar que el programa no identifica a la persona si se hace una nueva grabación, esto se debe a muchos factores como ya se habían mencionado al inicio de esta tesis.

Como se mencionó al inicio de la tesis una de las motivaciones fue entender la identificación de canciones que hacen algunos programas comerciales como Shazam, por lo que en esta tesis se hace el reconocimiento de canciones como ejemplo usando el software creado en esta tesis. A continuación se muestran algunos resultados.

Las canciones seleccionadas fueron Ares y Hades y la gráfica del archivo de sonido que se obtuvo fue el siguiente

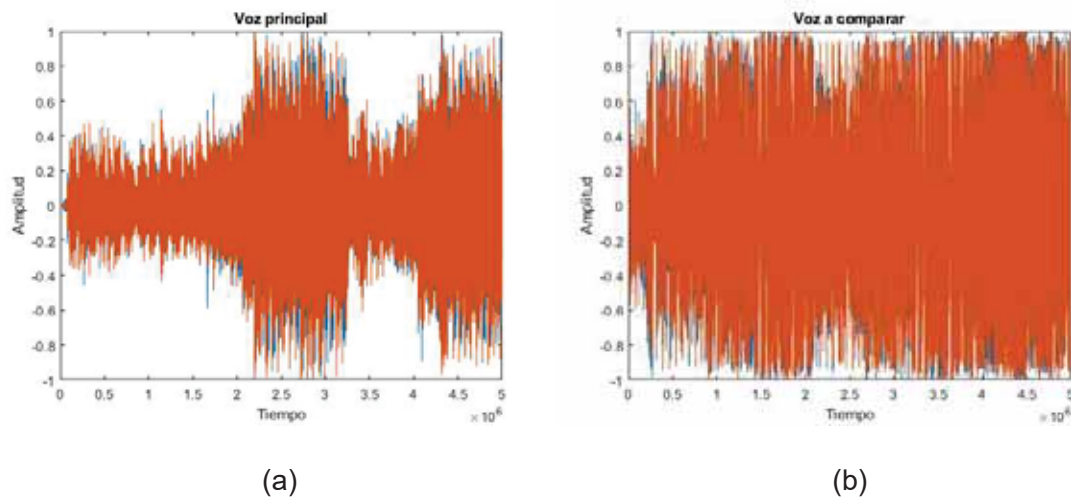


Figura 21: (a) Gráfica de la canción Ares que se va a reconocer, (b) Canción Hades que es con la que se va a comparar.

Para comparar La voz principal se utilizó la canción Hades, por lo que, de acuerdo al procedimiento planteado, las transformadas de Fourier que se utilizaron para este ejemplo se muestran en la figura 22.

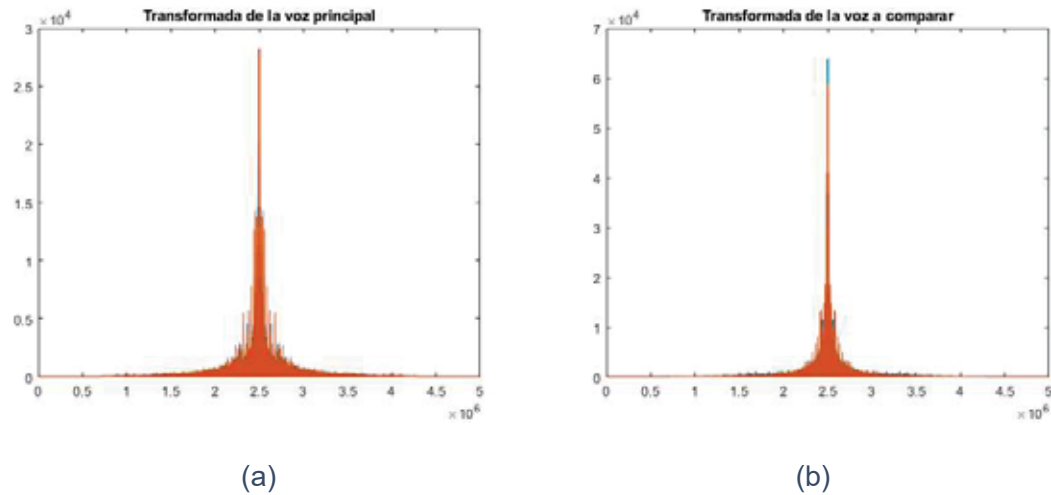


Figura 22: Transformada de fourier de las canciones que se usaron para el experimento.

De acuerdo al procedimiento de obtiene el conjugado de la canción Ares la cual ahora se llamará canción master o voz principal, por lo que ahora desempeñará el papel filtro.

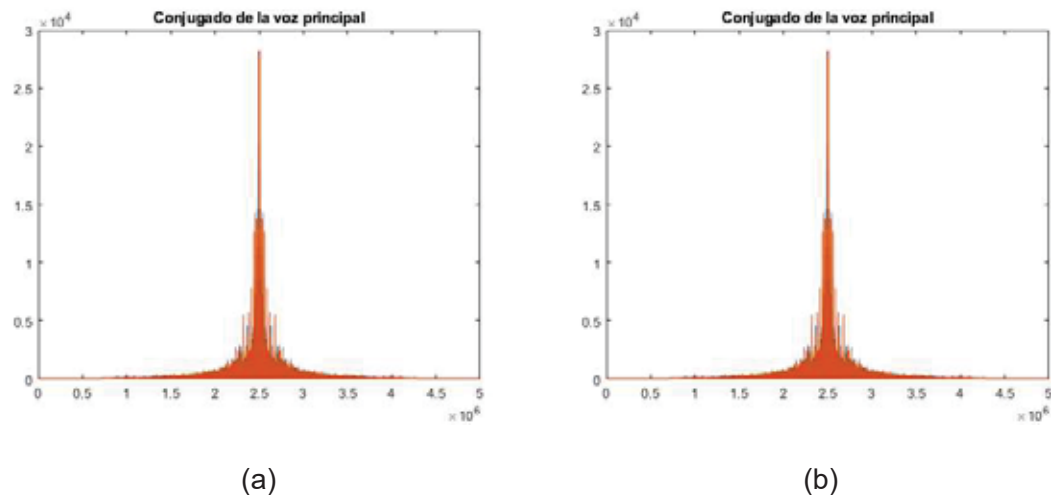


Figura 23: Se muestran las graficas del conjugado de la canción principal, que es la que se usará para mostrar los resultados.

En la figura 24 se muestra la multiplicación del filtro por la transformada de Fourier de la canción a comparar.

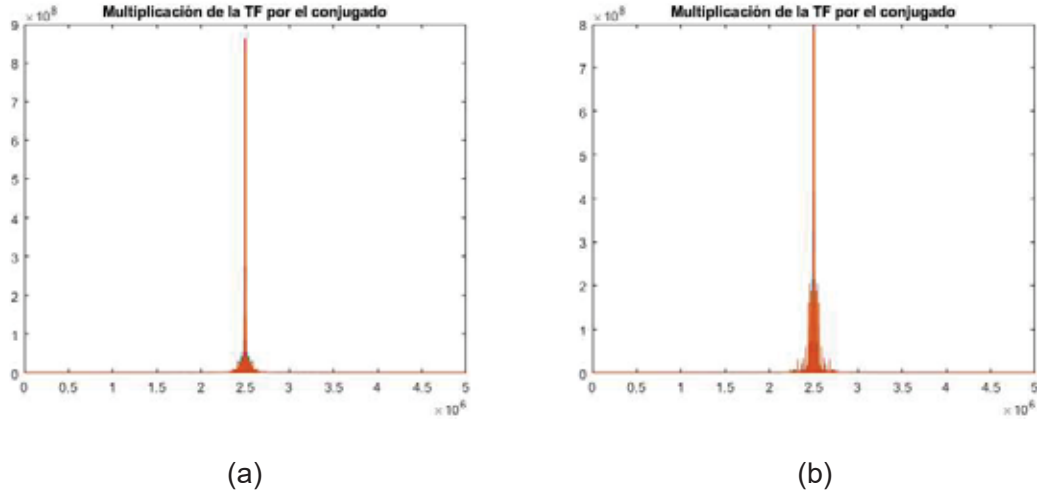


Figura 24: (a) La multiplicación del conjugado por la transformada de Fourier de la canción a comparar (b) la canción a comparar es la misma que la canción master.

En la figura 25 se muestra la transformada inversa de Fourier de la multiplicación que se hizo en la figura 24.

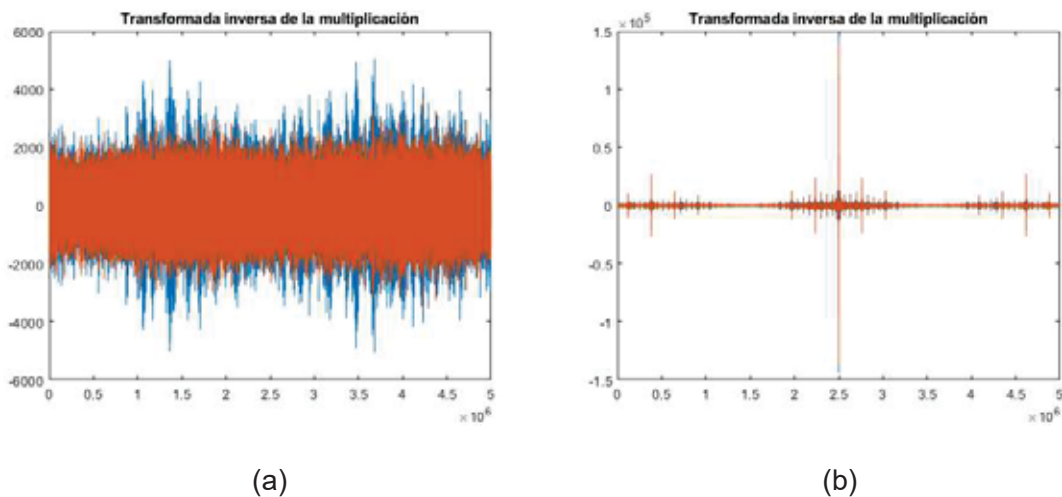
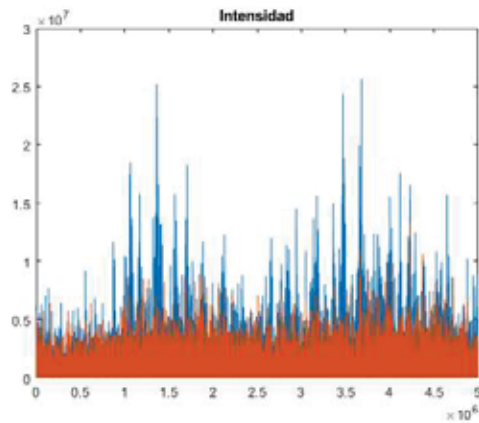
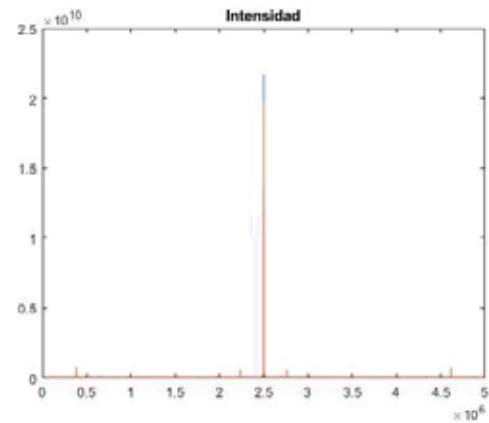


Figura 25: Se determina la transformada inversa de Fourier, (a) son las canciones diferentes, (b) son la misma canción.

En la figura 26 se muestra la intensidad, tal como se presentó con las voces en el ejemplo de las canciones se determinaron dos casos, uno en donde las canciones no coinciden (figura 26(a)) y en otro lado cuando las canciones coinciden (figura 26(b)). Se puede apreciar a simple vista la función delta que se está buscando.



(a)



(b)

Figura 26: Al determinar la intensidad en la figura (a) no es visible la delta o pico que se esta buscando por otro lado en (b) se es muy visible la dicha delta.

Capítulo VI

Conclusiones

En los sistemas de reconocimiento de voz hay muchas variables que pueden afectar dicho reconocimiento, la mayoría van desde las emociones con las que se dicen las palabras al momento de guardarlas, la edad, algún acento, incluyendo los problemas del habla como pueden ser la disfasia, dislalia, disfemia, tartamudez por mencionar algunos.

Pero con el reconocimiento de patrones que proporciona la transformada de Fourier por medio del conjugado, se puede plantear otro método de reconocimiento de voz, con el único detalle que se observó es que las muestras no eran reconocidas si se volvían a grabar de nuevo el archivo de audio, por lo que se intentó probar con canciones para ver si se podía observar otro resultado, resultando que si podía reconocer los patrones proporcionados por la transformada de Fourier al igual si se hacía una grabación de las mismas. Por lo que se puede determinar es que este sistema de reconocimiento de voz funciona para canciones, pero no para personas debido a los problemas ya antes mencionados.

Una cosa que se tiene que mencionar es que este código se hizo sin la idea de incluir algún método estadístico o de red neuronal esto debido a que se podían salir de los límites de esta investigación, debido a que la mayoría de programas trabajan con dichos aspectos.

Por lo que puede concluir final y satisfactoriamente que la transformada de Fourier cumple con las condiciones para crear un sistema de identificador de voz, ya que facilita mucho la obtención del espectro de voz.

Referencias

- Báez, L. D. (2006). *MATLAB con aplicaciones a la ingeniería, física y finanzas*. México: Alfaomega.
- Goodman, J. W. (2005). *Introduction to Fourier optics*. U.S.A: McGraw Hill.
- Hsu, H. P. (1987). *Análisis de Fourier*. U.S.A.: Schawm.
- Miyara, F. (2003). *Acústica y sistemas de sonido*. Rosario, Argentina: Universidad Nacional de Rosario.
- Óscar Reinoso García, L. M. (2018). *MATLAB: conceptos básicos y descripción gráfica*. Elche, España: Universidad Miguel Hernández de Elche.
- Robert, R., Halliday, D., & Krane, K. S. (2002). *Física Vol. 1*. México: Grupo Editorial Patria.
- Rodríguez, M. &. (2014). *Audiología*. España: Sociedad Española de Otorrinolaringología y Patología Cérvico-Facial.
- Rodríguez, R. &. (2006). *Manual de audioprotesismo*. Guadalajara, México: Blauton México.
- Smith, W. F., & Hashemi, J. (2006). *Fundamentos de la ciencia e ingeniería de materiales*. México: The McGraw-Hill.
- Valenzuela Ramírez, M. A. (2015). *Construcción de un mecanismo de procesamiento de patrones temporales aplicado al reconocimiento de voz*. Santiago, Chile: Universidad de Chile.
- Zemansky, M. W. (1985). *Calor y termodinámica*. México: McGraw-Hill.