

# Reconocimiento de voz mediante filtros digitales y técnicas de procesamiento digital de señales

Rodriguez Rivera A.M., Cortés Ramírez F., Valencia C., Sotelo S.

**Resumen** – En el presente trabajo se presenta una aproximación inicial a un proyecto que consiste en diseñar un sistema de reconocimiento de voz basado en filtros digitales y apoyado en técnicas de procesamiento digital de señales, con el fin de reconocer una palabra predefinida y activar con ello un LED (O una secuencia de LEDS) o un servo-motor.

**Índice de Términos** – Filtro digital, Voz, Señal, Función de transferencia, Transformada de Fourier

## I. INTRODUCCIÓN

Los filtros en general, son elementos que permiten clasificar señales eléctricas en función de su frecuencia. La versión más simple de un filtro se encuentra en los filtros pasa-bajas y pasa-altas, los cuales cuyo nombre indica, permiten pasar señales cuyas frecuencias sean mayores, o menores a una cierta frecuencia de corte, respectivamente. Dicha frecuencia de corte está determinada por los valores del condensador y de la resistencia puestos en el circuito. Si bien, los filtros análogos son funcionales, una variante, que a menudo posee mayor precisión, y capacidades para procesar la señal de entrada por un menor costo en componentes, son los filtros digitales, cuyo principio es recoger los datos de una señal y pasarlos por un algoritmo cuyo fin es alterarlos de tal forma que cumplan las características del filtro que se busca.

En ese orden de ideas, es posible diseñar un sistema basado en la idea de filtro digital que permita reconocer y/o clasificar palabras en función de parámetros como la frecuencia, usando de por medio técnicas de procesamiento digital de señales que permitan procesar de forma eficiente las señales de ejemplo y las señales de entrada

## II. FILTROS DIGITALES

### A. Definición

Un filtro digital, fundamentalmente es una operación matemática, que toma los valores de una señal de entrada y los transforma en un conjunto de valores de salida con ciertas características.

Un filtro responde a la señal de entrada con base en un parámetro específico, que es la frecuencia.

Un filtro se caracteriza por tres formas que son equivalentes entre sí, la respuesta al impulso, la respuesta en frecuencia y la respuesta al escalón.

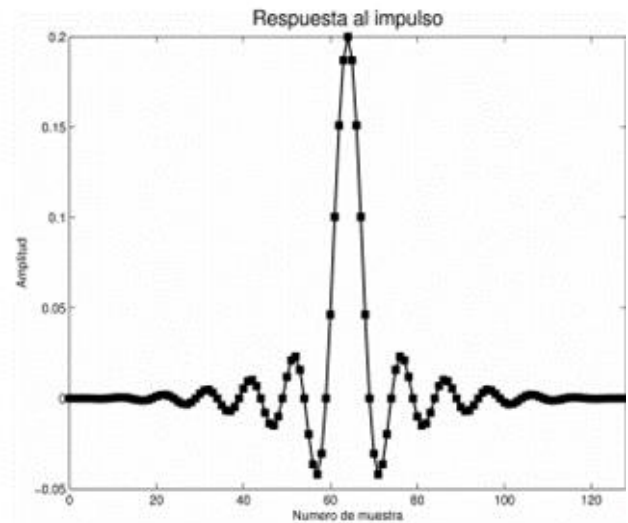


Figura 1. Respuesta al impulso de un filtro

Conociendo la respuesta al impulso, se puede calcular la respuesta del filtro a cualquier entrada (principio de superposición).

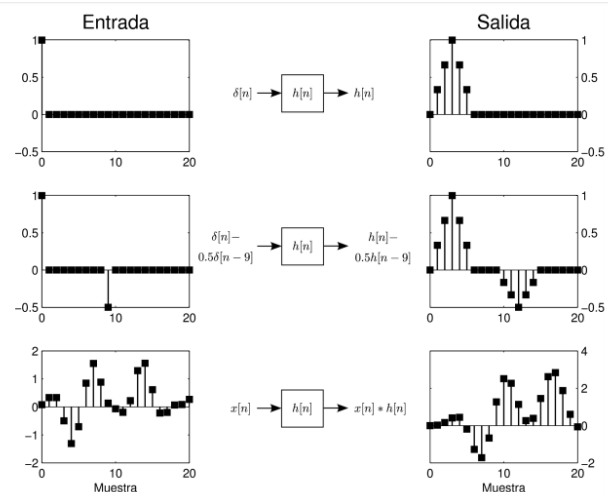


Figura 2. Ejemplo de superposición

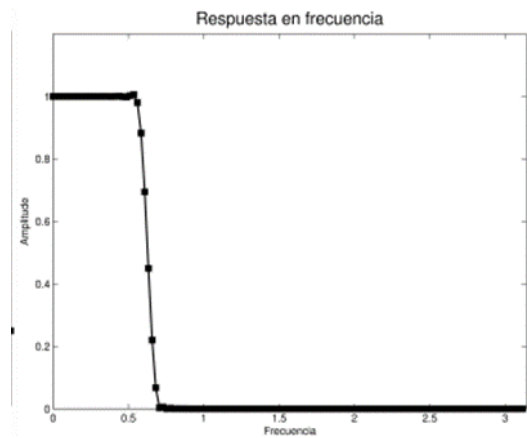


Figura 3. Respuesta en frecuencia a un filtro

Respuesta en frecuencia es la Transformada de Fourier, la cual más adelante explicaremos con más detenimiento.

Observaciones que hay que tener en cuenta:

- Generalmente es una función que toma valores complejos.
  - El periodo es de  $2\pi$ .
  - Al ser una función compleja, la podemos representar con parte real y parte imaginaria, o en su notación polar como magnitud y fase.
  - Se recomienda usar la notación polar ya que muestra la magnitud y la fase, con estos dos parámetros se pueden graficar en función de la frecuencia, generando así un diagrama de Bode el cual nos sirve para extraer información importante.

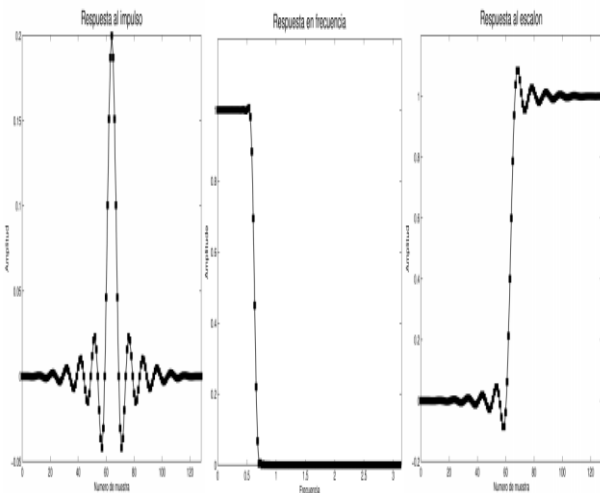


Figura 4. Respuesta al escalón de un filtro

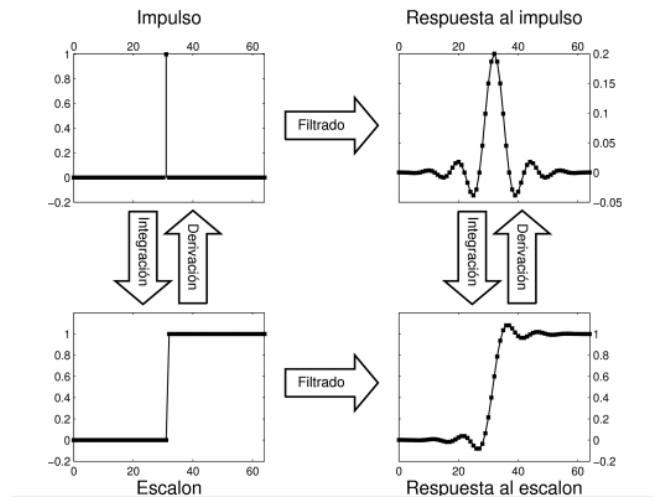


Figura 5. Respuesta al escalón de un filtro

En la figura 5 podemos ver la equivalencia entre la respuesta al impulso y respuesta al escalón: el escalón obtiene mediante la integración discreta del impulso.

En particular, es necesario fijarse en la respuesta en frecuencia del filtro, no obstante, las señales de voz en el entorno digital a menudo se obtienen en función del tiempo, por lo cual, será necesario transformarlas al dominio de la frecuencia, mediante una herramienta matemática, que es la Transformada de Fourier.

### B. Transformada de Fourier

Matemáticamente una transformada de Fourier puede definirse como:

$$X(\omega) \triangleq \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$$

Donde  $X(\omega)$  representa el contenido frecuencial de la señal  $x(n)$ .

En pocas palabras, la transformada de Fourier, permite transformar una señal en el dominio del tiempo al dominio de la frecuencia, que es precisamente el dominio en el que ocurrirá el procesamiento que se hará de la señal.

La transformada de Fourier es una herramienta matemática muy utilizada en los campos como la ingeniería y la acústica ya que permite como previamente se mencionó, representar una señal en el dominio de la frecuencia sin alterar o perder información de la señal original. Lo importante de este método es la posibilidad de por decirlo de alguna manera descomponer una señal compleja, en varias señales de menor complejidad con una única frecuencia. Hay que tener en cuenta que esta transformada se puede aplicar únicamente a señales estacionarias, o, dicho de otra forma, que sus parámetros sean constantes en el tiempo, lo que implica que si uno ve la señal en

$t=0s$  o  $t= 200s$  en esencia se verían igual.

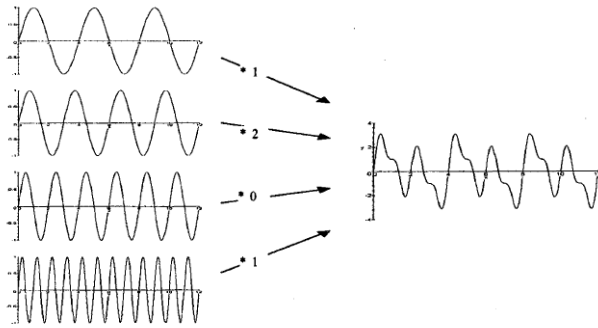


Figura 2. División de una señal compleja en un conjunto de señales mas sencillo.

La transformada de Fourier posee unas propiedades que nos limitamos a enunciar para el conocimiento del lector.

-Linealidad. Si las funciones  $x(t)$  y  $y(t)$  tienen como transformada a  $X(f)$  y  $Y(f)$ , respectivamente, entonces la suma de ambas tiene como transformada a  $X(f) + Y(f)$ .

-Simetría. Si  $X(f)$  es la transformada de  $x(t)$ , la transformada de  $X(t)$  es  $x(-f)$ .

-Escala en el tiempo y en la frecuencia. Si realizamos un escala de la variable  $t$  mediante la constante  $k$  la transformada da:  $\frac{1}{|k|} x\left(\frac{t}{k}\right) \leftrightarrow x(kf)$ .

-Desplazamiento en tiempo y en frecuencia. Si desplazamos el tiempo según la constante  $t_0$ , la transformada nos queda:

$$x(t - t_0) \leftrightarrow x(f)e^{-2\pi j f t_0}$$

## TIPOS DE FILTROS

Cuando hablamos de filtros hacemos referencia a algún sistema el cual discrimina entre lo que deseamos que pase y lo que no, dependiendo de que se quiera ver o escuchar, así las cosas, tenemos diversos tipos de filtros, entre los cuales los podemos caracterizar de la siguiente forma:

- Filtro pasabajas
- Filtro Pasaaltas
- Filtros pasatodo
- Filtro pasabandas

El primero tal como su nombre lo dice corresponde a un filtro el cual permite únicamente el paso de frecuencias bajas atenuando las frecuencias altas.

El segundo filtro de forma equivalente al primero, el nombre dice que corresponde a un filtro el cual atenúa las frecuencias bajas y en el caso de las frecuencias altas puede llegar a amplificarlas. Por otra parte, el considerar una frecuencia alta o baja va a depender del campo en el que estemos trabajando de esta forma si de habla de un filtro como en este caso para la voz, puede oscilar entre 250 Hz y 3000Hz.

El tercer tipo de filtro que encontramos es el filtro pasa todo, pero la primera impresión que se tiene al escuchar este filtro es

que permite el paso de todas las frecuencias, por lo que en sentido estricto no sería un filtro. Este tipo de filtro cubren todo el espectro frecuencial y su función es ofrecer una ganancia constante a todas las señales en cualquier frecuencia, sin embargo, la relación de la fase de la señal de entrada con respecto a la señal de salida va a variar como una función de la frecuencia.

Por último, tenemos el filtro pasa bandas, el cual es el que utilizaremos para realizar el proyecto. Este filtro delimita un rango de frecuencias comprendido entre la frecuencia de corte inferior y la frecuencia de corte:

- $F_{c1}$ : Frecuencia de corte del **filtro pasa alto**. (Frecuencia de corte inferior)

- $F_{c2}$ : Frecuencia de corte del **filtro pasa bajo**. (frecuencia de corte superior).

Naturalmente si se modifican los parámetros de los capacitores o los resistores del circuito naturalmente se va a modificar la las frecuencias y nuestro ancho de banda será mas grande o más pequeño:

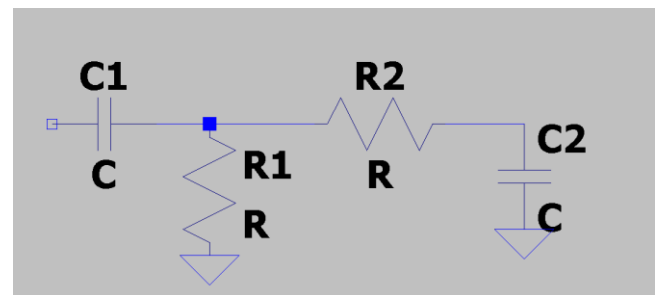


Figura 5. Circuito pasa bandas

Para este circuito si modificamos el resistor 1 y la capacitor 1, estaremos modificando el filtro para alta lo que quiere decir que podemos aumentar o disminuir el ancho de banda en su parte superior, por otra parte si modificamos el resistor 2 y el capacitor 2 estaremos modificando el filtro pasa baja, de igual forma disminuyendo o aumentando su rango frecuencial.

Esto se diseña de tal forma que, como se ve en el circuito, es para que el segundo se comporte como una carga para el primero y en lo posible se espera que este primero, sea lo mas bajo posible, de tal forma que el segundo filtro le demande muy poca corriente al primero.

De esta forma con el segundo arreglo a una frecuencia de corte mas alta, se asume que los valores de impedancias que son generadas por  $R_2$  y  $C_2$  sean altos, de tal forma que esto sea una carga menor para el primer filtro.

## RESPUESTA EN FRECUENCIA DE UN FILTRO RC PASABANDA

Como se puede ver en la figura número 5, el filtro 1, compuesto por el capacitor 1 y el resistor 1 permite el paso de frecuencias superiores a la frecuencia de corte, la cual viene

determinada por el valor que se le asigne a la tanto a  $R$  como a  $C$ . Las ondas que logren superar este filtro pasaran inmediatamente por el siguiente filtro, el número dos, el cual de igual forma permitira el paso unicamente de frecuencias superiores a la frecuencia de corte determinada en este caso por el capacitor 2 y el resisor 2.

Hay otro escenario en el cual tanto  $R_1$  como  $R_2$  y  $C_1$  como  $C_2$  son iguales, lo que implica que la frecuencia de corte sean las mismas, lo que a su vez conlleva que la unica frecuencia que pase por este circuito sea la frecuencia de corte.

Por otra parte también esta el escenario en el cual la frecuencia de corte filtro pasa altas es mayor que la frecuencia de corte del filtro pasa bajo, lo que conlleva que aca no pase ninguna frecuencia, ya que la ventana de frecuencias esta invertida.

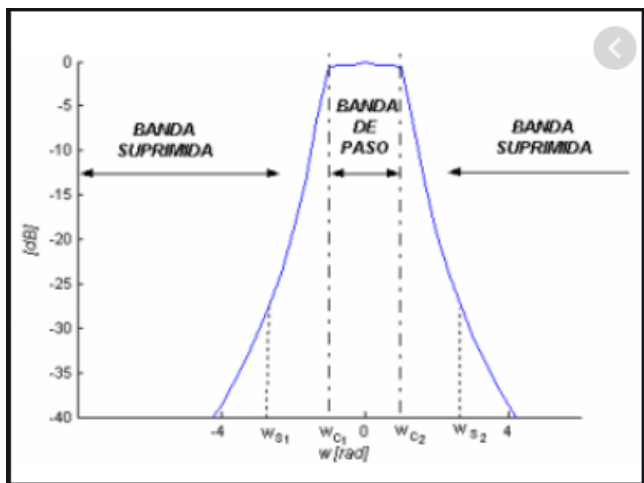


Figura 6. Ejemplo de filtro pasa bandas

La respuesta del filtro pasa banda que se puede ver en la figura 6, resulta de la combinación, de como anteriormente se menciono los dos filtros individuales (filtro pasa altas, filtro pasa bajas).

En el escenario en el cual la frecuencia de corte sea equivalente para ambos filtros, el retardo de base del filtro pasa bajo se cancelará con el efecto en adelante del filtro pasa alto por lo tanto la única frecuencia sin desfase será la de la frecuencia de corte.

## APARATO FONADOR HUMANO

En este apartado se pretende hacer una breve explicación acerca de lo que en nuestro caso va a producir la señal a la que le haremos el procesamiento digital.

El aparato fonador o también conocido como aparato vocal, es la parte encargada de producir o emitir sonido.

Los órganos necesarios para este fin son: por el sistema respiratorio tenemos los pulmones, los bronquios, la laringe y la tráquea, por el aparato digestivo tenemos los dientes, labios, lengua y paladar, finalmente pero no menos necesario las cuerdas vocales.

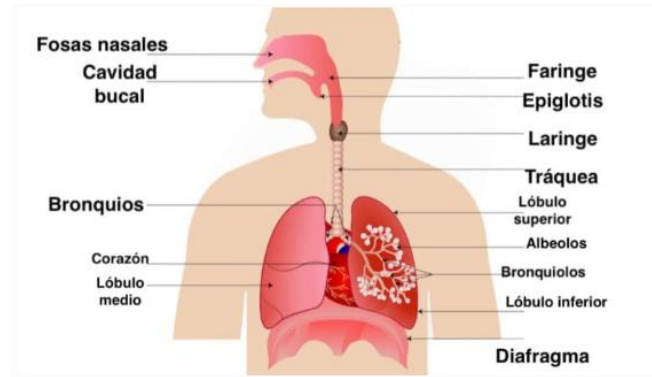


Figura 7 Ilustración del aparato fonador

Los pasos para el que el aparato fonador funcione son los siguientes:

En primer lugar, los pulmones se llenan de aire, con la presión que ejerce el diafragma, se desocupan empujando este aire hacia la boca por la tráquea. El aire con ese impulso se encuentra con las cuerdas vocales, que al vibrar llenan el aire de ondas sonoras.

En segundo lugar, el aire ya sonorizado alcanza la laringe y la faringe, donde en vez de ser exhalado por la nariz, se dirige hacia la boca donde es modulado.

En tercer lugar, este aire ya sonorizado llega a la boca, y es liberado hacia el exterior luego de que los órganos de la boca se haya colocad o en alguna posición deseada para generar algún sonido específico, ya sea abriendo o cerrando la cavidad bucal, posicionando la lengua en el camino del aire o haciéndolo rebotar e distintas partes del paladar.

La función del aparato fonador es la de producir sonidos, los cuales pueden ser articulados, transformados en una cadena hablada, lo que implica poder decir palabras.

## PROCESAMIENTO DIGITAL DE SEÑALES

### Conversión Analógica-Digital

#### Proceso de cuantificación:

La información digital difiere de su correspondiente representación analógica en dos aspectos:

Está muestreada  
Esta cuantificada

Para manejar información en un computador o en algún hardware, es necesario cual información se debe mantener y cual se puede perder sin que esta perdida afecte la aplicación o lo que se desea hacer.

Para esto se necesita definir:

Frecuencia de muestreo de la señal.

Número de bits a utilizarse para representar cada dato.

Tipo de filtro analógico que se requiere para convertir de analógico a digital.

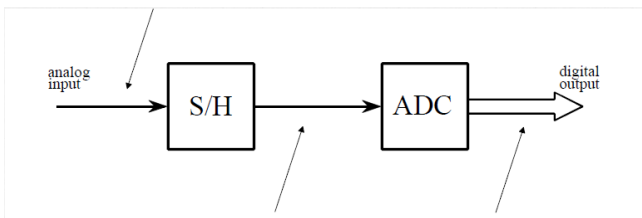


Figura 10 Proceso de conversión A-D

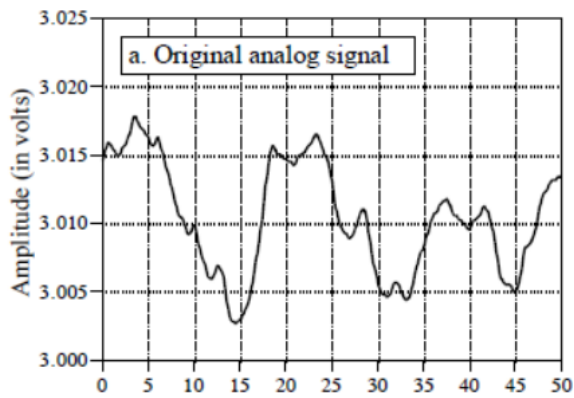


Figura 11 Ejemplo de señal analógica

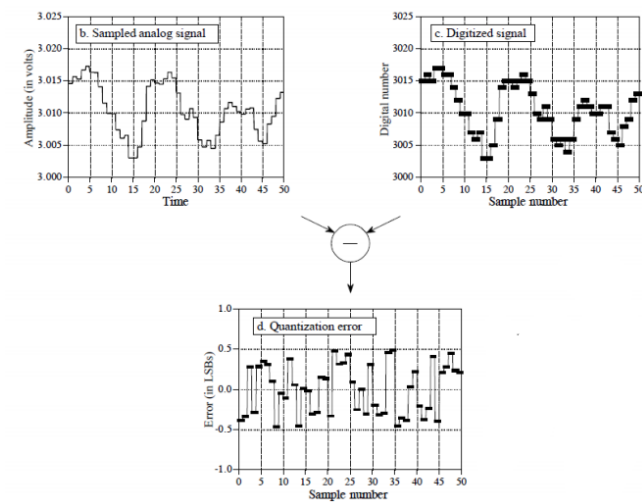


Figura 12 Señal analógica y señal digital

Para el ejemplo anterior podemos ver el voltaje está entre 0 V y 4.95 V.

Se pretende usar 12 bits para representar los datos en forma digital; entonces existen  $2^{12} = 4096$  posibles valores, que puede representar desde el valor 0 hasta el valor 4095.

El cambio entre los valores de entrada y los muestreados (figura a y b) ocurre en tiempos específicos y uniformes. En esta parte del proceso, la variable independiente (tiempo) se convierte de continua a discreta.

Durante la segunda parte del proceso que convierte los datos muestreados y cuantizados (figura b y c), a cada valor en b se

le asigna su correspondiente valor en alguno de los representados por los bits.

Lo que implica una aproximación. Por ejemplo, 2.56001 y 2.5600 se convertirán al mismo valor: 2,560. Aquí la variable dependiente (voltaje) se convierte de continua a discreta.

### TEOREMA DE MUESTREO

Un muestreo correcto es aquel que genera puntos suficientes para reconstruir la señal original, sin ninguna ambigüedad.

El teorema de muestreo de Shannon, conocido también como razón de muestreo de Shannon, establece la relación entre la frecuencia de la función con la frecuencia de muestreo lo que permite un muestreo adecuado.

### FRECUENCIA DE MUESTREO

Es una función continua  $x(t)$ , la variable continua  $t$  se reemplaza con  $nT_s$ , que es un valor discreto de dicha variable  $t$  continua.

La variable  $n$  (discreta) representa el subíndice del arreglo de datos y  $T_s$  es el periodo de muestreo, esto es, el tiempo que hay entre dos muestras.

$$f_s = \frac{1}{T_s}$$

Figura 12 Señal analógica y señal digital

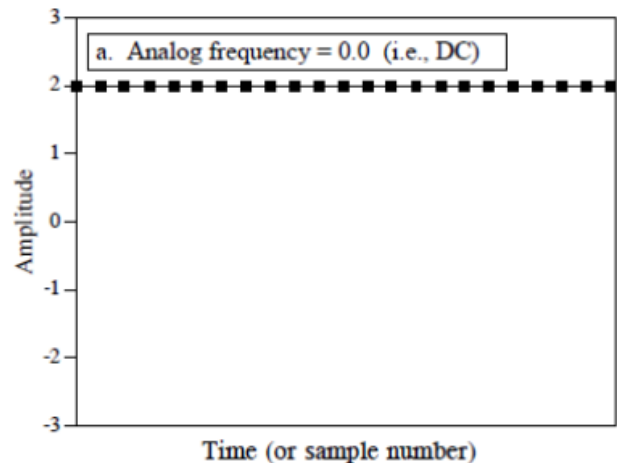


Figura 13 Ejemplo de un muestreo adecuado

Señal con DC constante: función coseno con frecuencia 0.



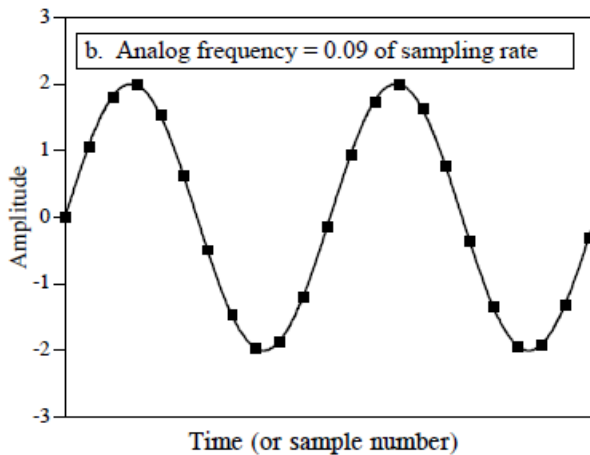


Figura 14 Ejemplo de un muestreo adecuado

Función seno con frecuencia 90 ciclos/Seg, muestreada a 1000 muestras /seg.

Del ejemplo anterior  
 $f = 90$  ciclos/ segundo

$$f_s = \frac{1000 \text{ muestras}}{\text{segundo}}$$

$$F = 0.09 f_s$$

$$F = (0.09)(1000) = 90$$

$$f_s = f / 0.09$$

$$F(9/100) = 1.1 F$$

#### SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB

[4] El sistema de reconocimiento de voz permite que los usuarios graben palabras por medio de un micrófono y que esta sea reconocida en la base de datos que se genere previamente. El sistema posee un entorno gráfico en la computadora, que proporciona las selecciones de grabación, donde la señal de voz es ingresada a la computadora y es procesada por los algoritmos del programa que modifican la señal, obteniendo los parámetros significativos de la señal de voz, para luego ser almacenados en la computadora. La selección de reconocimiento permite que la palabra sea comparada con la base de datos almacenada en la computadora, dicha base de datos y ya fue procesada digitalmente por el programa. Esta selección reconoce la palabra.

Este entorno gráfico proporciona, por otra parte, un análisis gráfico de las palabras grabadas y reconocidas. Como el sistema de Reconocimiento de Voz es un sistema de procesamiento digital de señales de voz, el análisis gráfico en el reconocimiento es un análisis del espectro de frecuencias de la señal de voz.

El procesamiento digital de señales de voz tiene una gran variedad de aplicaciones, existe una base para el tratamiento

digital de señales, que puede ser implementada para lograr obtener lo que nos interese según la aplicación.

El sistema de Reconocimiento de voz es una de las aplicaciones del procesamiento digital de señales de voz. El sistema consiste en obtener una señal de voz que permita reconocer qué palabra se está hablando. Consta de una interfaz gráfica que permite la interacción del usuario por medio de un micrófono con la computadora, la que procesa automáticamente los datos adquiridos. Basado en los resultados de este sistema, se puede ver cómo se plantea la base del procesamiento digital de señales de voz y queda a la libre imaginación cómo puede ser utilizado para otras aplicaciones, además del de reconocimiento de voz.

#### Componentes del sistema

Micrófono

Fidelidad

Directividad

Ruido de fondo

Rango dinámico

Impedancia interna

Factor de directividad

#### FORMANTES

Los formantes son elementos que sirven para distinguir componentes del habla humana, principalmente, las vocales y sonidos sonantes. El formante con la frecuencia más baja se llama F1, en el segundo F2, y el tercero F3, etc. Normalmente sólo los dos primeros son necesarios para caracterizar más formantes. Generalmente, los formantes posteriores determinan propiedades acústicas como el timbre.

Los dos primeros formantes se determinan principalmente por la posición de la lengua. Sucediendo que F1 que tiene una frecuencia más alta cuanto más baja está la lengua, es decir una mayor abertura. Para el F2 tiene mayor frecuencia cuanto más hacia delante está posicionada la lengua.

No todos los sonidos se componen de formantes definidos. Solamente aparecen en sonantes, que incluyen los sonidos pulmonares: vocales, aproximantes y nasales. Los nasales tienen un formante adicional F#, en torno a los 1500 Hz.

Si la frecuencia fundamental es mayor que la frecuencia de los formantes, entonces el carácter del sonido se pierde y se vuelven difíciles de distinguir, por lo cual son difíciles de reconocer.[4]

<b>Formantes Vocálicos</b>	
<b>Vocal</b>	<b>Región principal formantica</b>
/u/	200 a 400 Hz
/o/	400 a 600 Hz
/a/	800 a 1200 Hz
/e/	400 a 600 y 2200 a 2600 Hz
/i/	200 a 400 y 3000 a 3500 Hz

Figura 15 Formatos vocálicos

## REFERENCIAS

- [1] Gómez. Juan Carlos, “Teoría de sistemas y señales”
- [2] Psenicka, Bohumil , “*Procesamiento digital de señales*”.
- [3] Semillero de procesamiento de señales e Imágenes PROMISE, UR- Escuela Colombiana de Ingeniería Julio Garavito
- [4] G. Velasquez,” SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB”, Trabajo de grado, Universidad De San Carlos De Guatemala, Guatemala, 2008.
- [5] “Procesamiento digital de señales Semana 2. (Primera parte) Conversión Analógica-Digital”-Coordinación de computación, INAOE, 2017
- [6] J. Máxima (16/05/2020). Aparato Fonador [online]. Disponible en: <https://www.caracteristicas.co/aparato-fonador/>
- [7] “Filtro RC Pasa Banda- Filtro Pasa band a RC” Disponible en : <https://unicrom.com/filtro-rc-pasa-banda/>