RECONOCIMIENTO DE PATRONES AUDITIVOS EN AMBIENTES RUIDOSOS

BELMER ALBERTO CORDOBA DIAZ

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
BOGOTÁ
2009 II

RECONOCIMIENTO DE PATRONES AUDITIVOS EN AMBIENTES RUIDOSOS

BELMER ALBERTO CORDOBA DIAZ
031026
betuncio32@hotmail.com

MONOGRAFÍA DE GRADO

ASESOR TÉCNICO
ING. DIEGO DÍAZ MUÑOZ
DSP (PROCESAMIENTO DIGITAL DE SEÑALES)

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
BOGOTÁ
2009 II

Nota de aceptación

_____
_____
_____
_____
_____
_____

_____
Ing. Diego Díaz Muñoz
Asesor

_____
Ing. Freud Romero
Jurado 1

_____
Ing. Rafael Cubillos
Jurado 2

Bogotá (06, Noviembre, 2009)

A mis padres por el amor y consejos
brindados durante toda mi vida,
además de su apoyo incondicional.

AGRADECIMIENTOS

Ante todo gracias a Dios, por iluminarme el camino del éxito y acompañarme en él, por darme la valentía y fortaleza de seguir perseverando en la realización de todos mis sueños.

A mis padres Belmer Cordoba C. y María Helena Diaz H. y mi hermana Mary Gelen Cordoba D., quienes siempre han estado ahí, brindándome sus consejos, apoyo incondicional y amor que me sirven de fortaleza para seguir adelante, por contribuir en mi formación personal y hacer de mí la persona que soy.

A mis abuelos Marcial Alberto Cordoba y José Antonio Diaz por haberme inculcado desde niño el amor a los libros, el estudio y el trabajo fuerte, pero quizás la enseñanza más importante que me pudieron dejar fue su ejemplar vida ya que sus consejos me hicieron el hombre que soy ahora.

A mi abuelita Celinda Cabrera de Cordoba por ser una apoyo incondicional en todos los momentos difíciles que he tenido en mi vida, por creer siempre en mis capacidades y por formarme como una persona de bien.

A mi abuelita Graciela Herrera de Diaz que a pesar de permanecer a mi lado en este mundo poco tiempo se que desde el cielo aporto más de un granito de arena para este logro tan importante en mi vida.

A mi tía Celia Cordoba C. quien creyó en mí sin importar las adversidades que se me presentaron. ¡Muchas Gracias tía, te debo una!.

A mi tía Susana Velasco H. quien es una persona ejemplar de vida y quien me acogió como uno más de sus hijos.

A Cindy Tatiana Chitiva P. quien durante un largo tiempo se ha convertido en una persona muy importante en mi vida brindándome la estabilidad emocional y su apoyo incondicional.

A Juan David Prieto. y Eliana Marcela García H. quienes contribuyeron con sus conocimientos, interés y apoyo para que este proyecto saliera adelante.

Al ingeniero Diego Díaz, quien ayudo a encontrar respuestas a mis dudas, por brindarme sus consejos, ánimos, conocimientos y aportar en mi formación como ingeniero.

A Manuel Alejandro Jiménez V. por sus consejos y ánimos durante los últimos años de carrera, le deseo el mejor de los éxitos en su vida profesional.

A la Universidad, sus ingenieros, profesores y administrativos que brindaron sus conocimientos, ayuda, servicio e instalaciones para el desarrollo de mis estudios.

A cada una de las personas que contribuyeron directa o indirectamente en la realización de este proyecto.

CONTENIDO

LISTA DE TABLAS

pág.

LISTA DE FIGURAS

pág.

LISTA DE ANEXOS

RESUMEN

Este proyecto busca generar una aplicación que reconozca patrones de voz basándose en un algoritmo de filtrado Mel, la cual es una herramienta matemática del análisis de señales de audio que puede implementarse para el reconocimiento de palabras y fonemas.

Para obtener un buen reconocimiento de patrones de audio se deben seguir los siguientes pasos:

- Digitalización de la señal de audio. Durante este proceso la señal pasa del dominio del tiempo continuo al dominio del tiempo discreto, por esto se utilizará un archivo de audio con extensión .wav, el cual es generado por la grabación realizada.
- El archivo de audio debe ser filtrado. Implementando un filtro digital equiripple, generando un nuevo archivo de audio con un ancho de banda de 8KHz.
- Se le aplica la transformada discreta de Fourier, esto con el fin de representar la señal de audio como su espectro en el dominio de la frecuencia.
- El análisis sinusoidal de la voz. Por medio de este se simboliza las características más relevantes de la señal de audio plasmadas en un espectrograma de frecuencias.
- Representación de la energía en ciertas frecuencias a lo largo del tiempo, esto para la observación y análisis del espectrograma de la señal de audio tanto en entorno ruidoso como en entorno sin ruido.
- Realización de la comparación de los espectrogramas, con el fin de corroborar la existencia o no de la palabra o fonema a reconocer.

INTRODUCCIÓN

En los sistemas de reconocimiento de voz no se intenta reconocer el sonido del fonema, sino identificar una serie de características principales para saber si el locutor dijo lo que se presume. El tamaño de la “frase” en el reconocimiento de voz afecta su complejidad.

El comportamiento de los sistemas de reconocimiento del habla se degrada rápidamente debido a la presencia del ruido de fondo, recientemente se ha propuesto una técnica de representación de la señal de voz basada en predicción lineal. Que ha mostrado ser atractiva para el reconocimiento de señales de audio en condiciones severas de ruido gracias a su simplicidad computacional.

El problema del reconocimiento de voz permanece sin resolver aun en caso de palabras aisladas y vocabularios pequeños. Por esta razón se ha propuesto diversas técnicas de reducción de ruido en cada una de las etapas del proceso de reconocimiento especialmente en extracción de parámetros fonéticos y medidas de similitud.

La etapa de parametrización es dada por un código de predicción lineal. Usado ampliamente en reconocimiento de sonidos fonéticos, este código es sensible al ruido blanco, pero aun así esta técnica es favorable respecto a otras técnicas de reconocimiento auditivos como lo son, bancos de filtros y vector de cuantización. Con esto la técnica de predicción lineal en combinación con el procedimiento matemático de Autocorrelación permite una aproximación a un buen proceso de reconocimiento de patrones de audio.

El procedimiento de desarrollo del proyecto se llevará a cabo inicialmente con un estudio de las técnicas de reconocimiento observando ventajas que pueda tener alguna sobre los demás, además si se pueden mezclar para obtener mejores resultados en el proceso de reconocimiento.

1. PROBLEMA

En el entorno se define como ruido todo sonido no deseado por el receptor, además en el ámbito de la comunicación es aquel que no contiene información clara o que el receptor no es capaz de identificar o comprender. Entonces en el ambiente ruidoso existen perturbaciones que sufre la señal en el proceso comunicativo y muchas veces se busca individualizar o separar la señal de la respectiva perturbación.

En realidad, la mayoría de las ondas son el resultado de muchas perturbaciones sucesivas del medio y no solo de una, produciendo sonidos aperiódicos, es decir que las sucesivas perturbaciones se producen a intervalos irregulares evitando mantener una constante forma de onda, aclarando que en el medio ambiente una señal acústica no puede ser representada como un sonido periódico.

Investigar sobre la detección de señales acústicas particulares en entornos ruidosos es importante, ya que con esta, se demuestra que las señales de audio a pesar de las pérdidas en sus características, es posible desarrollar métodos de detección de ciertos patrones en particular, el cual es el primer paso para la reconstrucción de la señal original.

Dentro del reconocimiento de patrones auditivos y los problemas que se encuentran al realizarlo se tiene que es muy importante relacionar los campos que intervienen en el mismo, como lo son, el procesamiento de la señal, física (acústica), Patrón de reconocimiento, teoría de la información y comunicación, lingüística, fisiología, informática, sicología (Rabiner, 1993). Luego de haber identificado los campos que deben ser aplicados en uno o más problemas de reconocimiento de patrones auditivos, se obtienen que las principales aplicaciones de este tema se relacionen con seguridad e identificación de tracto vocal.

Desarrollar un reconocimiento de patrones de audio en ambientes donde no sea fácil la identificación del mismo, esto con el fin de darle una gran cantidad de aplicaciones. Algunas preguntas claves en reconocimiento de digital de fonemas son ¿Cómo son comparados los patrones fonéticos? y ¿Cómo determinar las similitudes entre ellos?.

2. JUSTIFICACIÓN

Dependiendo de la posición de las diferentes articulaciones del tracto vocal y nasal se pueden presentar una serie de problemas los cuales son conocidos como ortognáticos y estos pueden ser identificados a través de los sonidos que se producen en la voz y que posteriormente pueden ser plasmados en espectrogramas de frecuencia.

Al mencionar un problema ortognáticos se dice que este es una deformación o maloclución esquelética debido a una mala posición de los dientes o una deformación de sus bases óseas respecto al resto de la cara. Por medio del uso de plantillas se pueden identificar problemas ortognáticos gracias a simulaciones acústicas realizadas al individuo y con esto poder referenciar las deformaciones craneales que presente cada persona. Vale la pena mencionar que los sonidos de los dinosaurios fueron posibles de identificar gracias a simulaciones acústicas sobre los cráneos de estos grandes animales.

Los avances tecnológicos en reconocimiento auditivos también buscan un desarrollo favorable para el hallazgo de pistas que puedan involucrar a un sospechoso tanto en una escena previa como en la ejecución del crimen, dando así herramientas de trabajo para los entes gubernamentales que imparten justicia en la solución de múltiples crímenes que muchas veces son truncados por la falta de pistas que puedan incriminar a un delincuente.

En el campo de la seguridad también puede implementarse el reconocimiento de patrones auditivos en el uso de firmas digitales fortaleciendo el tema de seguridad para así evitar el aumento de hurtos de objetos valiosos o simplemente como elemento que impida el acceso a un lugar determinado de personal no autorizado.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Seleccionar e implementar un algoritmo para el reconocimiento de patrones de audio en el espectro de la voz en entornos ruidosos.

3.2 OBJETIVOS ESPECÍFICOS

- Reseñar los métodos de reconocimiento de patrones para muestras de audio en el espectro de la voz.
- Determinar los parámetros bajo los cuales se puede realizar el reconocimiento de patrones de audio en el espectro de la voz para implementar un algoritmo.
- Determinar un algoritmo fácilmente implementable para el reconocimiento de patrones de audio en el espectro de la voz.
- Implementar el algoritmo para el reconocimiento de patrones de audio en el espectro de la voz.
- Realizar una interfaz gráfica que permita mostrar los valores obtenidos para su análisis.
- Seleccionar un patrón de audio en el espectro de la voz para reconocerlo bajo parámetros definidos.

4. MARCO REFERENCIAL

4.1 ANTECEDENTES

Desde los años treinta del siglo XX donde se inicio el estudio del reconocimiento de la voz, hasta la actualidad se han realizado muchos avances tecnológicos para que cada vez más el habla comience a ser habitual en un número cada vez mayor de personas (Marti, 1996).

Durante más de treinta años los avances tecnológicos eran publicados solo en congresos científicos. Solo esporádicamente la opinión pública recibía información de los avances por medio de los medios de comunicación. En la década de los ochenta cuando la reunión entre distintas disciplinas permite la utilización práctica de los sistemas de reconocimiento de voz. En primer lugar se utilizaron los modelos ocultos de **Markov** (empleados por primera vez en por el Istitute for Defense Analyses de IBM y por Dragon System) los cuales permitieron mejorar el modelado de las características y propiedades de los sonidos. Este cambio implicó el avance en la forma de enfrentarse a un problema de reconocimiento de voz al pasar de la tecnologías de reconocimiento de patrones (como Dynamic Time Warping) a las basadas en modelo estadístico, que son las que se siguen empleando en la actualidad. En segundo lugar el desarrollo de ordenadores más potentes dotados con tarjetas de procesamiento de señal, permitió la implementación de reconocedores que funcionen en tiempo real. Por último, el sistema de almacenamiento masivo de datos junto con el gran esfuerzo para realizar bases de datos de voz, permitió que empezara a disponer de grandes cantidades de voz para poder estimar adecuadamente los parámetros estadísticos. En este sentido cabe resaltar el esfuerzo realizado por las organizaciones **LDC** (Linguistic Data Consortium), en Estados Unidos y **ELRA** (European Lenguage Resources Association) en Europa (Marti, 1996).

Cabe destacar que ya en la época de los sesenta aparecieron esquemas como el TANGORA, de IBM, que era dependiente del locutor y permitía dictar pequeños informes. Así mismo, en esta década en los laboratorios de Bell de AT&T, se comienza a trabajar en sistemas que hicieran posible un reconocimiento fuera independiente del locutor (Marti, 1996).

A finales de los años ochenta, la organización Defense Advance Research Projects Agency (**DARPA**) de Estados Unidos invirtió en un programa de investigación orientado al reconocimiento del habla continua hasta 1000 palabras. Desde entonces, DARPA ha continuado con el patrocinio a investigaciones en el área del habla continua para vocabularios cada vez mayores, hasta llegar al momento actual donde se trabaja en reconocimiento del habla espontanea para vocabularios ilimitados (Marti, 1996).

Con todo lo expuesto se aprecia que los avances conseguidos en estos cincuenta años han sido impresionantes, donde hemos pasado de los reconocedores de sonidos aislados dependientes del locutor a sistemas de reconocimiento del habla espontanea para vocabularios ilimitados. Sin embargo, a pesar de todos los avances conseguidos, aun quedan muchos problemas tecnológicos por resolver, que aun hacen que los reconocedores sean parte débil de los sistemas conversacionales o del dialogo (Marti, 1996).

La problemática del reconocimiento está relacionada específicamente con la falta de robustez de estos sistemas frente a la variabilidad del mundo real. Esto, mientras que en las respuestas de laboratorio las respuestas son bastante buenas, cuando la tecnología de reconocimiento es aplicada a aplicaciones reales en las que las condiciones del laboratorio ya no se mantienen constantes, las tasas de error aumentan. Es por tanto, en la última década, cuando se toma conciencia de la complejidad e importancia de las técnicas que permiten que un sistema con estas características se adapte automáticamente a las situaciones de su entorno. Las fuentes de variabilidad que afectan a un sistema de reconocimiento son el canal, el ruido de fondo, la variabilidad inter/intralocutor, el cambio del dominio semántico, y las características del habla espontanea (Marti, 1996).

4.2 MARCO CONCEPTUAL

4.2.1 TEORÍA DE FILTROS

En la teoría de filtros el principal término a tener en cuenta es la definición de función de transferencia y su notación está dada por $\mathrm{H}(s)$. La función de transferencia es la relación de la magnitud de la señal de salida en transformada de Laplace sobre la magnitud de la señal de entrada en transformada de Laplace con condiciones iníciales nulas como se ve en la Ecuación 1:

Ecuación 1

$$H(s) = \frac{V_{out}(s) - V_{out}(0)}{V_{in}(S) - V_{in}(0)}$$

Ahora la función de transferencia es utilizada para expresar la respuesta en frecuencia. La transformada de Laplace se realiza para pasar del dominio del tiempo al dominio de la frecuencia, donde en este último se ofrece una expresión general para el dominio de la variable compleja (Gabiola, 2007).

Por medio de la función de transferencia y de la respuesta en frecuencia se puede determinar la definición de filtros y los tipos de filtros existentes. La primera parte dice que un filtro es un cuadripolo cuya respuesta en frecuencia se adecua a ciertas especificaciones. La respuesta en frecuencia tiene dos partes, la respuesta en amplitud y la respuesta en fase (Gabiola, 2007).

Filtro pasa bajos: es aquel que mantiene una amplitud constante hasta determinada frecuencia, después de esta frecuencia su magnitud vale 0. Por ende solo deja pasar las bajas frecuencias (Gabiola, 2007).

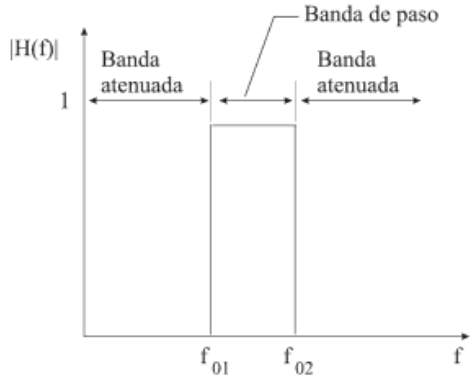
Filtro pasa altos: al contrario del filtro pasa bajos este mantiene el valor de 0 hasta cierta frecuencia, después de esta frecuencia mantiene un valor constante en amplitud. Con esto solo deja pasar las frecuencias altas (Gabiola, 2007).

Filtro pasa banda: idealmente es un filtro que toma el valor de 0 para todas las frecuencias excepto para un rango de frecuencias conocido como ancho de banda en el cual este toma un valor de amplitud constante (Gabiola, 2007).

Filtro rechaza banda: al contrario del filtro pasa banda este filtro toma un valor de amplitud constante para todas las frecuencias excepto un ancho de banda determinado donde este toma un valor de amplitud de 0 (Gabiola, 2007).

En los filtros existen rangos de frecuencia que dicho en otras palabras pueden denominarse como bandas de paso y bandas atenuadas. Las bandas de paso es aquel intervalo de frecuencias donde la respuesta en frecuencia toma un valor constante. Mientras que las bandas atenuadas es un rango de frecuencias donde la respuesta en frecuencia toma un valor en amplitud nula como se muestra en la Figura1 (Cadavid, 2004).

Figura 1. Banda de paso y banda de rechazo

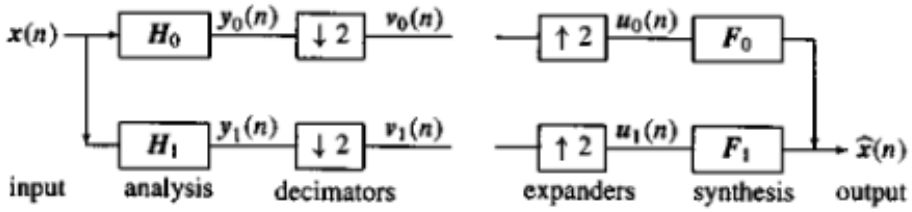


(Cadavid, 2004)

4.2.2 BANCO DE FILTROS

Los bancos de filtros son un conjunto de filtros utilizados en reconstrucción perfecta de señales, los cuales están vinculados por muestreo de operadores y algunas veces por retardos. El bajo muestreo de los operadores es reducido en décadas, mientras que los altos muestreos son expandidos. En un banco de filtros de dos canales, el análisis de los filtros es pasa bajos o pasa altos respectivamente. Esos son los filtros H_0 y H_1, donde se hace el banco de análisis como se muestra en la Figura 2.

Figura 2. Banco de filtros (Análisis y Síntesis)



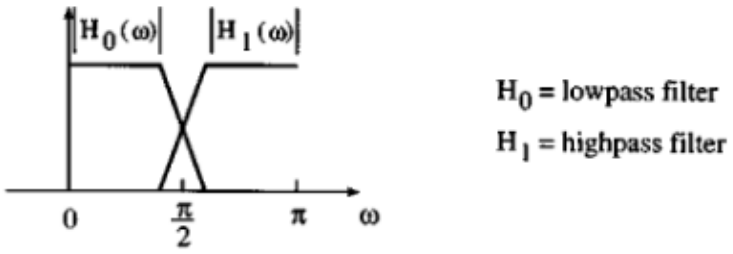
(Truong, 1996)

Luego de escoger correctamente los filtros H_0 y H_1 se debe diseñar F_0 y F_1 para obtener una reconstrucción perfecta de la señal. La brecha de la Figura 2 indica donde los bajos muestreos podrían ser codificados para su almacenamiento o

transmisión en el punto donde es posible que la señal se comprima o se destruya. En la reconstrucción perfecta se asume que no existe la compresión, por lo que la brecha es estrecha (Truong, 1996).

Cuando se habla de H_0 se indica que es el filtro pasa bajos y cuando por el contrario se menciona a H_1 se dice que este indica que este es el filtro pasa altos. Un boceto de la respuesta en frecuencia generalmente esta dado por la Figura 3 donde muestra que estos filtros no son ideales y donde las respuestas se sobreponen. Y donde se puede presentar distorsión de amplitud y distorsión de fase (en la Figura 3 no se muestra la fase).

Figura 3. Filtro pasa altos y pasa bajos



(Truong, 1996)

Cuando se realiza la síntesis de los filtros F_0 y F_1 es necesario que estos deban ser especialmente adaptados al análisis de los filtros H_0 y H_1. Con el fin de evitar errores que puedan presentarse en este banco de análisis. Cuando se aplica un banco de filtros se tiene como primer objetivo una reconstrucción perfecta de una señal. Para esto se necesita que el banco de filtros sea biortogonal. Dentro del proceso de biortogonalidad se debe tener claro que el banco de análisis debe ser inverso al banco de síntesis (Truong, 1996).

4.2.3 SEÑALES BANDA BASE

El sistema de transmisión banda base analógico es un sistema de transmisión que no posee un desplazamiento en frecuencia o mejor dicho no posee modulación alguna.

Existen dos reglas para el realización de su estudio ya que son muy pocos los sistemas que trabajan en banda base. Los sistemas de modulación pueden estudiarse como banda base y muchos de los conceptos y parámetros de banda base pueden ser aplicables a los primeros. El sistema banda base sirve como elemento de comparación de los atributos de los distintos tipos de modulación (Fernandez Rubio, 1999).

En los sistemas de transmisión banda base se requieren una serie de elementos los cuales son la fuente, el canal de transmisión y el receptor. Junto con esto se deben tener en cuenta ciertos términos que hacen parte de cada uno de los elementos del sistema de transmisión banda base conocidos como potencia de la señal transmitida, potencia de la señal recibida, potencia de la señal en destino y potencia de ruido en destino (Fernandez Rubio, 1999).

Fuente: Este elemento es aquel donde se genera el mensaje, vale aclarar que este mensaje puede ser cualquier señal. En caso de que la señal no sea eléctrica se requiere de un transductor para poder ser interpretada (Fernandez Rubio, 1999).

Transmisor: Aquí la señal mensaje es transmitida a través del canal de comunicaciones. El transmisor en banda base simplemente se amplifica para proporcionar la potencia necesaria para la señal y posteriormente ser recepcionada. Para un sistema modulado requerirá su propio sistema de modulación (Fernandez Rubio, 1999).

Canal de comunicaciones: Este canal se representa básicamente con una representación geométrica de una línea de transmisión esto solo a nivel banda base pero cuando se implementa un sistema de modulación se requieren otros tipos de medios como lo son fibras ópticas, guías de ondas con frecuencia en el orden de microondas además de medios no guiados. Conforme se realiza la propagación se presenta atenuación en la señal esto se debe al efecto Joule en conductores, pérdida de potencia radiada en el espacio libre estas pérdidas se caracterizan por medio de ausencia de distorsión (Fernandez Rubio, 1999).

Receptor: Puede denominarse de una forma simple como el elemento capaz de recuperar el mensaje, en este punto es cuando se presentan problemas ya que no solo se afecta el mensaje enviado con problemas de atenuación sino que es afectada con problemas de ruido y distorsión (Fernandez Rubio, 1999).

Para finalizar con el receptor es necesario saber que en un sistema sin distorsión y con ruido blanco aditivo, posee un relación señal a ruido y está dada por la potencia transmitida, por las perdidas en el canal, por la densidad espectral de potencia de ruido y por el ancho de banda del receptor, además de anotar que es independiente de la ganancia del receptor (Fernandez Rubio, 1999).

En la Tabla 1 se muestran unos ejemplos de relaciones señal a ruido en algunos sistemas de transmisión banda base analógico. Algo más para agregar es que cuando el ruido no es aditivo la relación señal a ruido es confusa y no tiene gran significado.

Tabla 1. Relaciones señal a ruido en algunos sistemas de Tx

TIPO DE SEÑAL	BANDA DE FRECUENCIA	S/N dBs
Voz inteligible	500Hz - 2KHz	5 - 10
Voz calidad telefónica	300Hz – 3.4KHz	25 - 35
Radiodifusión AM	100Hz – 5KHz	40 - 50
Alta fidelidad audio	20HZ – 20KHz	45 - 65
Televisión video	60Hz – 4.2KHz	45 - 55

(Fernandez Rubio, 1999)

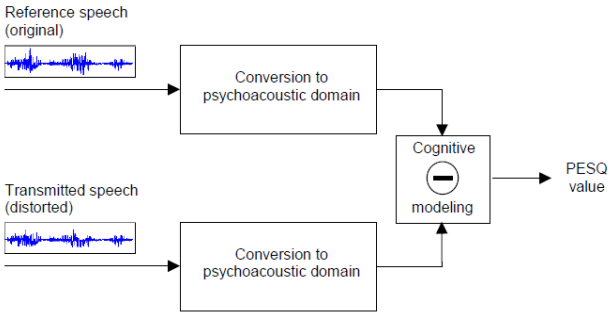
Lo necesario en el uso de sistemas de transmisión banda base es el uso de repetidores los cuales se caracterizan por ser filtros mas amplificadores además se encuentran entre el transmisor y el receptor (Fernandez Rubio, 1999).

4.2.4 ALGORITMO PESQ

Antes de saber cuál es el algoritmo PESQ se debe saber que este es utilizado para realizar una evaluación perceptual de la calidad de voz, además es un método objetivo para la estimación de la calidad objetiva de redes móviles. PESQ es un método par que la gente experimente la calidad de voz en una red móvil. El propósito de PESQ es imitar la percepción de los sonidos de los humanos. En él se evalúa la calidad de una señal de voz distorsionada en comparación con la original sin distorsiones de señal (AB, 2006).

En PESQ, la señal original y la señal distorsionada de voz se asignan en representaciones psicofísicas que coinciden con la forma en que los humanos interpretan la voz. La calidad de la voz distorsionada se evalúa sobre la base de las representaciones psicofísicas. En la Figura 4 se observará una descripción general del algoritmo PESQ (AB, 2006).

Figura 4. Descripción general algoritmo PESQ

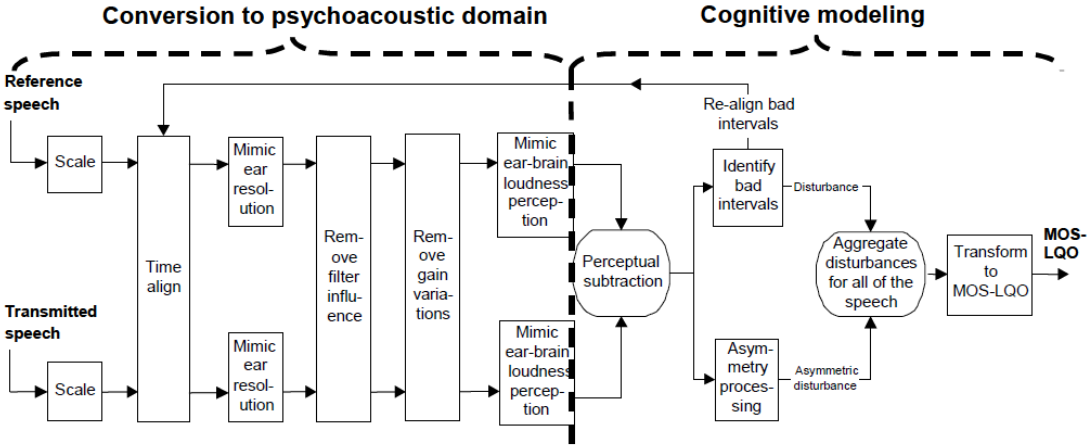


(AB, 2006)

El algoritmo PESQ produce valores de respuesta de 1 a 4.5. Un valor de 4.5 significa que la medida de voz no sufrió ningún tipo de distorsión, mientras que un valor de 1 significa que la medida de la voz sufrió una grave degradación (AB, 2006).

El algoritmo PESQ consiste de dos partes, la primera se encarga de una conversión al dominio psicoacústico y la segunda de un modelamiento cognitivo. Los pasos más importantes en cada parte se ven en la figura 5.

Figura 5. Partes del algoritmo PESQ



(AB, 2006)

Escala: tanto lo transmitido y la referencia de voz son escaladas para compensar el aumento de la red (AB, 2006).

Alineamiento de tiempo: el tiempo de la voz de referencia y el tiempo de la voz de transmisión son alineados para que coincida todas las partes entre los dos (AB, 2006).

Imitar la resolución del oído: transformar la señal de voz al dominio de la frecuencia esto con el fin de tratar de imitar el oído en la forma que trata diferentes frecuencias (AB, 2006).

Remover la influencia del filtro: Como su nombre lo dice este paso busca eliminar el efecto del filtrado, esto se hace utilizando la medición en la función de transferencia del filtrado para ecualizar la señal de voz de referencia (AB, 2006).

Eliminar las variaciones de ganancia: control automático de ganancia para ya que los elementos de red pueden provocar variaciones de ganancia (AB, 2006).

Imitar la percepción sonora oído-cerebro: interpretación de la intensidad del espectro en que el oído humano transforma la intensidad sonora percibida (AB, 2006).

Sustracción perceptiva: las señales se restan teniendo en cuenta como el cerebro percibe las diferencias (AB, 2006).

Identificar intervalos erróneos: se presentan gracias a una alineación incorrecta para intervalos de la voz de referencia (AB, 2006).

Procesamiento de asimetría: si un códec de voz añade ruido a la señal de voz original, una distorsión claramente audible, para esto se calcula la asimetría de la densidad de la señal de perturbación (AB, 2006).

Agregar perturbaciones para todas las frecuencias: primero que todo son sumadas perturbaciones en la frecuencia plana. Este resulta en perturbación y perturbaciones de señales asimétricas las cuales son representadas como distorsiones de voz durante periodos de tiempo muy corto (AB, 2006).

Transformar a MOS-LQO: se representa el valor que dice que tanta degradación sufrió la señal de voz trasmitida con la que se tomo de referencia (AB, 2006).

4.3 MARCO TEÓRICO

4.3.1 PRODUCCION DE LA VOZ

La voz es una mezcla de sonidos y se realizan a través de un proceso donde se intervienen el tracto vocal y el tracto nasal. El primero está compuesto por la apertura de las cuerdas vocales y finaliza en los labios, y consiste en la conexión del esófago con la boca donde el área seccional cruzada del tracto vocal determina en qué posición están la lengua, la mandíbula, los dientes y los labios. El segundo inicia en el velo y finaliza en la nariz. Cuando el velo (mecanismo situado en la parte posterior de la cavidad vocal) es reducido, el tracto nasal esta acústicamente acoplado al tracto vocal y produce el sonido nasal de la voz (Rabiner, 1993).

Cuando el flujo del aire es expelido por los pulmones a través de la tráquea, la elasticidad de las cuerdas vocales dentro de la laringe vibra por el flujo del aire, el cual es fraccionado en pulsos casi periódicos los cuales luego son modulados en frecuencia y pasados a través de la faringe, la cavidad vocal y posiblemente la cavidad nasal. Dependiendo de la posición de diferentes articulaciones (mandíbula, labios, dientes, velo, boca) son producidos diferentes sonidos (Rabiner, 1993).

La señal de la voz es un señal variable lentamente en tiempo en el sentido que cuando examinamos sobre un periodo corto de tiempo que puede oscilar entre los 5 y 100ms estas características son bastante estacionarias, sin embargo sobre largos periodos de tiempo en el orden del medio segundo las características de la señal cambian a fin de reflejar los diferentes sonidos de la voz (Rabiner, 1993).

En la conversión de los sonidos de la voz en forma de señal se requieren de la discusión de algunas técnicas fundamentales usadas para proveer las características que se presentan en todo el sistema de reconocimiento de voz. En particular dos técnicas conocidas y métodos ampliamente usados de análisis de espectros. El primero llamado banco de filtros, enfocado en el método de predicción lineal. El segundo método, está basado en el procesamiento dentro del sistema auditivo humano (Rabiner, 1993).

4.3.2 MODELO DE CODIFICACIÓN DE PREDICCIÓN LINEAL

Dentro de la teoría de codificación de predicción lineal (LPC), se debe describir un modelo general a través de un proceso de reconocimiento de voz. Para que esto sea posible, debe conocerse el siguiente seguimiento (Rabiner, 1993):

- LPC provee un buen modelo de señal de voz, esto es dado gracias al estado de la región sonora de la voz en los cuales los modelos de LPC proveen una buena aproximación del tracto vocal (Rabiner, 1993).
- LPC es un modelo analíticamente manejable. Este método es matemáticamente preciso, simple y fácil de implementar en cualquier software o hardware (Rabiner, 1993).
- El modelo LPC trabaja bien en aplicaciones de reconocimiento. La experiencia ha mostrado que los reconocedores de voz basados en LPC son comparables o mejor que los reconocedores por bancos de filtros (Rabiner, 1993).

La idea básica del modelo LPC es que dada una muestra de voz en un tiempo n,$s(n)$, puede ser aproximada como una combinación lineal del pasado p de las muestras de voz, tal que:

Ecuación 2

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \cdots + a_p s(n-p)$$

Donde los coeficientes $a_1, a_2, \ldots, a_p$ son asumidos como constantes en el cuadro de análisis de voz. Con eso la Ecuación será representada junto con un término de excitación $Gu(n)$ dando:

Ecuación 3

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + Gu(n)$$

Donde $u(n)$ es una excitación normalizada y G es la ganancia de esa excitación. Para la Ecuación en el dominio de z se obtiene la relación:

Ecuación 4

$$S(z) = \sum_{i=1}^{p} a_i \, z^{-1} S(z) + GU(z)$$

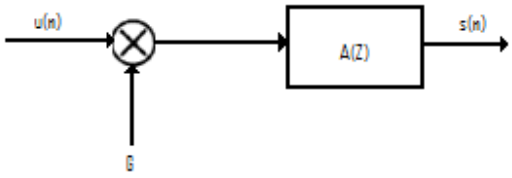
Llevando a la función de transferencia en la Ecuación 5:

Ecuación 5

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i \, z^{-1}} = \frac{1}{A(z)}$$

La representación de la función de transferencia está dada por la Figura 6**.**

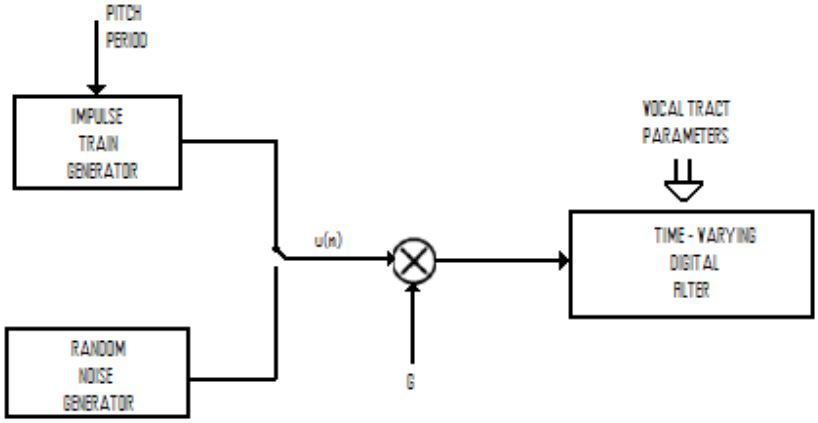
Figura 6. Representacion de la funcion de transferencia del analisis LPC



(Rabiner, 1993)

La síntesis apropiada del modelo para la voz, correspondiente al análisis LPC, que se muestra en la Figura 7. Aquí la fuente de excitación normalizada es seleccionada por un interruptor, en el cual la posición es controlada por un carácter sonoro/no sonoro de la voz, los cuales son escogidos con un tren de pulsos casi periódicos o una secuencia aleatoria de ruidos. La ganancia apropiada, G, es dada por la señal de voz, y la fuente escalada es utilizada como entrada a un filtro digital, los cuales son controlados por las características de los parámetros del tracto vocal siendo producidas por la voz (Rabiner, 1993).

Figura 7. Análisis LPC



(Rabiner, 1993)

4.3.3 VECTOR DE CUANTIZACIÓN

El resultado de cada análisis de banco de filtros y análisis LPC son una serie de vectores característicos de la variación espectral del tiempo de la señal de voz. Por conveniencia se denotan los vectores espectrales como v_l, $l = 1,2, \dots, L$ donde

cada vector es p-dimensional. Si se compara la proporción de información de la representación vectorial de la forma de onda sin codificar, se verá que el análisis espectral tuvo una reducción significativa de la proporción de información requerida (Rabiner, 1993).

A continuación se analizarán las ventajas y desventajas de este tipo de representación. Las ventajas de la representación de Vector de cuantización (VQ) son:

- Reduce el almacenamiento por análisis espectral de información.
- Reduce computación determinando la similitud del análisis espectral de vectores, debido a que esta se reduce a una tabla de look up de similitudes de vectores de libros de código.

Las desventajas del uso de VQ son:

- Existe una distorsión espectral inherente en representación del actual vector de análisis.
- El almacenamiento requerido para el libro de código de vectores no es despreciable.

4.3.4 PATRÓN Y TECNICAS DE COMPARACIÓN

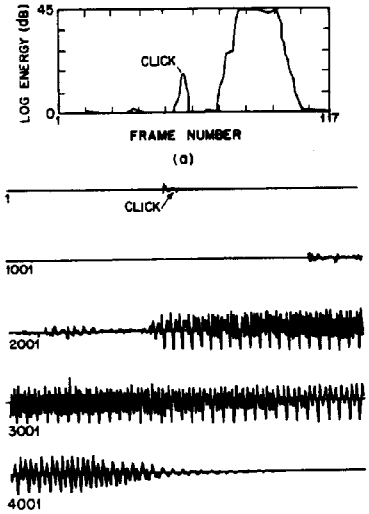
Una pregunta clave en reconocimiento de voz es como patrones de voz son comparados y determinar las similitudes entre ellos. Dependiendo de las especificaciones del sistema de reconocimiento, el patrón de comparación puede ser hecho en una ancha variedad de opciones (Rabiner, 1993).

El objetivo de la detección de voz es separar eventos acústicos de interés en una señal grabada continuamente de otra parte de la señal. La necesidad de detección de voz ocurre en muchas aplicaciones de telecomunicaciones. En sistemas de transmisión multicanal analógicos, una técnica llamada interpolación de voz por asignación de tiempo (TASI) es a menudo usada para tomar ventaja del canal de tiempo de ocio por detección de presencia del locutor y asignando un canal si usar solo cuando la voz es detectada para permitir más servicios a clientes que los canales normalmente proveen (Rabiner, 1993).

Otra pregunta clave en reconocimiento como con precisión la voz debe ser detectada para proveer el mejor patrón de voz por reconocimiento. En la Figura 8 se muestra un típico chasquido de boca producido por la apertura de los labios.

Cuando la boca está seca causando en la boca un pequeño estallido (Rabiner, 1993).

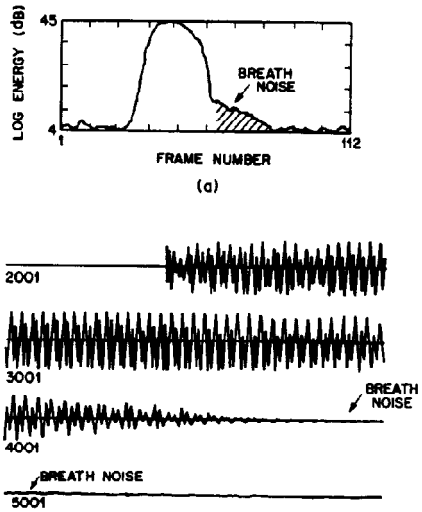
Figura 8. Ejemplo del procedimiento del chasquido de Boca



(Rabiner, 1993)

En la Figura 9 se muestra un nivel alto de ruido de la respiración producido al final de hablar causado por la respiración pesada del locutor. Esto típicamente ocurre cuando un locutor es de respiración corta y combina una respiración pesada con la conversación (Rabiner, 1993).

Figura 9. Muestra en el tiempo de un nivel alto de Respiración



(Rabiner, 1993)

4.3.5 ÓPTIMO RECONOCIMIENTO AUTOMÁTICO DE LA VOZ

El óptimo reconocimiento automático de voz tiene lugar donde las circunstancias son idénticas en las que el sistema de reconocimiento está capacitado. En las aplicaciones del mundo real esto nunca sucede. Hay variabilidad entre de las fuentes que producen desajustes entre la información y las condiciones de testeo. Dependiendo de su estado físico o emocional, un locutor produce sonidos con variaciones no deseadas entonces no se trasmite la información acústica pertinente. Técnicas robustas constituyen un área fundamental para el reconocimiento de la voz. Los retos actuales para el reconocimiento automático de la voz se enmarcan dentro de estas líneas de trabajo (Garcia Luz, 2008):

- El reconocimiento de códigos de voz a través de canales telefónicos. Esta tarea constituye una dificultad adicional y es que cada teléfono tiene su propio canal SNR y su respuesta en frecuencia. El reconocimiento de voz sobre líneas telefónicas debe realizar un canal de adaptación con algunos datos específicos de los canales (Garcia Luz, 2008).
- Bajo entornos SNR, hay diferentes campos donde se presenta el reconocimiento automático de voz y estos son:
- En teléfonos móviles
- Traslado de vehículos
- Discursos espontáneos
- Voz enmascarada por otras voces
- Voz enmascarada por música
- Ruidos no estacionarios
- Interferencia en un canal de voz. Las interferencias causadas por otras voces constituyen un desafío que los cambios en el medio ambiente debido al reconocimiento de toda la banda de ruidos (Garcia Luz, 2008).
- Rápida adaptación en habladores no nativos. Hay una actual demanda en reconocedores de voz para darle solidez y adaptación en los acentos para habladores no nativos (Garcia Luz, 2008).

Bases de datos con degradaciones realistas. Formulación, registro y difusión de la voz en bases de datos con ejemplos de degradación reales existentes en la práctica para hacerle frente a los retos existentes en el reconocimiento de voz (Garcia Luz, 2008).

4.3.6 EFECTOS DEL RUIDO ADITIVO

En el marco del reconocimiento automático de la voz, el fenómeno del ruido se puede definir como el sonido no deseado que distorsiona la información transmitida en la acústica señal dificultando su correcta percepción existen dos principales fuentes de distorsión de la señal de voz: ruido aditivo y canal de distorsión. La distorsión del canal se define como el ruido convolucional mezclado con la palabra en el dominio del tiempo. Al parecer como consecuencia de la transmisión de reverberaciones de la señal, la respuesta en frecuencia del micrófono utilizado o peculiaridades del canal de transmisión como en un filtro eléctrico. Los efectos del canal de distorsión se han librado con cierto éxito, ya que una vez convertida la señal en lineal estos efectos son analizados en el dominio de la frecuencia. Las técnicas tal como el filtrado RASTA, cancelación del eco o sustracción media cepstral han demostrado eliminar esos efectos de canal de distorsión (Cadavid, 2004).

El ruido aditivo es sumado a la señal de voz en el dominio del tiempo y en el dominio de la frecuencia no es fácil de remover ya que tiene la particularidad de transformar la voz no linealmente en determinados ámbitos del análisis. Hoy en día el ruido aditivo constituye la fuerza de conducción de la investigación del ASR (Reconocimiento Automático de la Voz): ruido blanco aditivo, Door Slams, superposición espontanea de voces y música de fondo, etc (Cadavid, 2004).

El modelo más utilizado para analizar los efectos del ruido en la comunicación oral representa el ruido como una combinación de aditivos y el ruido después de la expresión convolucional dada por:

Ecuación 6

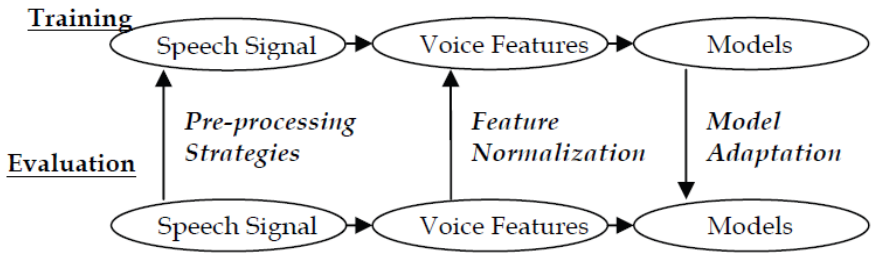
$$y[m] = x[m] * h[m] + n[m]$$

Suponiendo que el componente de ruido $n[m]$ y la señal de voz $x[m]$ son estadísticamente independientes, el resultado es la señal de voz ruidosa $y[m]$ (Cadavid, 2004).

4.3.7 TECNICAS DE RECONOCIMIENTO DE VOZ ROBUSTO

Existen varias clasificaciones de las técnicas existentes para el reconocimiento de voz robusto frente a los cambios del medio ambiente y del ruido. Una clasificación comúnmente utilizada es la que divide las técnicas procesamiento previo, las técnicas de normalización y las técnicas del modelo de adaptación de acuerdo al punto del proceso de reconocimiento como se observa en la figura 10 que muestra las estrategias de reconocimiento robusto (Garcia Luz, 2008).

Figura 10. Estrategias de reconocimiento robusto



(Garcia Luz, 2008)

- Técnicas de procesamiento previo de la señal: su objetivo es eliminar el ruido de la señal de voz antes de la parametrización con el fin de obtener una señal parametrizada limpia de ruido. Se basan en la idea de que la voz y el ruido están sin correlacionar y por lo tanto estas se suman en el dominio del tiempo. En consecuencia los espectros de potencia de una señal ruido será la suma de los espectros de potencia de la voz y el ruido. Las principales técnicas que se pueden desarrollar en este grupo son sustracción espectral lineal, sustracción espectral no lineal, filtrado de Wiener o la regla de supresión de ruido de Efraín Malah (Garcia Luz, 2008).
- Características en las técnicas de normalización: en el medio ambiente una vez que se elimina la distorsión de la señal de voz que ha sido parametrizada. A través de diferentes técnicas como el filtrado Cepstral Pasa altos, modelos de efectos de ruido, recuperan las características de voz limpia de las características de voz ruidosa (Garcia Luz, 2008). Se pueden encontrar cuatro sub-categorías dentro de este grupo de técnicas las cuales son:
- Técnicas de filtrado de paso de bandas altas: Añaden un alto nivel de solidez al reconocimiento con un bajo costo.
- Compensación de ruido con datos estéreo: este grupo de técnicas las características del ruido se la voz con los datos de estéreo limpios.
- Compensación de ruido basado en un modelo de entorno: estas técnicas dan una expresión analítica de la degradación del entorno.
- Algoritmos de igualamiento estadístico: Conjunto de algoritmos para la normalización de características que definen las transformaciones lineales y no lineales con el fin de modificar las estadísticas del ruido de la voz y hacerlos iguales a los de la voz limpia.

Técnicas del modelo de adaptación: estas técnicas hacen lo posible por hacer que el proceso de clasificación de las características de voz ruidosa sea óptimo. Los modelos acústicos obtenidos durante la fase de training son adaptados a las condiciones de prueba usando un conjunto de datos de adaptación del entorno ruidoso. Este procedimiento es usado tanto para la adaptación del entorno como para la adaptación de la persona que habla. Las más comunes estrategias de adaptación son MLLR (Regresión Lineal de Probabilidad Máxima), MAP (Adaptación A-posteriori Máxima), PMC (Combinación del Modelo Paralelo) y transformaciones del modelo no lineal como algunos interpretes usados en redes neurales (Garcia Luz, 2008).

4.3.8 ESTIMACIÓN DE LA EXPRESION DE INFORMACIÓN EN EL ESPECTRO DE LA VOZ

Los sonidos de la voz son producidos por el paso de una señal a través de un filtro del tracto vocal. La producción de los sonidos de la voz es asociada con la vibración de las cuerdas vocales. Debido a esto, la fuente de la señal consiste en la repetición periódica de pulsos y su espectro se aproxima a una línea consistente del espectro de la frecuencia fundamental y sus múltiplos (denominadas armónicas). Como resultado el procesamiento a corto plazo, el espectro de Fourier de corto tiempo de un segmento de voz expresado puede ser representado como una suma de escala (en amplitud y frecuencia) pasando por las versiones de la transformada de Fourier de la función Frame-window (ventaneo por trama). La estimación de la expresión de carácter de una región de frecuencia puede ser realizada basándose sobre la comparación del espectro de magnitud de corto tiempo de la señal con el espectro de la función Frame-window, el cual es el inicio del algoritmo de estimación de expresión. Este algoritmo no requiere información de la frecuencia fundamental. Sin embargo, si se dispone de esa información puede ser incorporado dentro del algoritmo (Garcia Luz, 2008).

El algoritmo de estimación de la expresión de información en el espectro de la voz viene descrito por una serie de pasos dados a continuación:

- Cálculo del espectro de la magnitud a corto plazo: un Frame (trama) de señal en el dominio del tiempo es ponderado por una función de ventaneo de análisis de Frame, expresada por ceros y la FFT (transformada de Fourier rápida) es aplicada para proporcionar un espectro de magnitud de corto tiempo. Durante toda la descripción del algoritmo se trabajará en la muestra de señales con $\mathrm{F_s} = 8\mathrm{kHz}$, la longitud del Frame de 256 muestras y la longitud de la FFT de 512 muestras (Garcia Luz, 2008).
- Cálculo de la Distancia de Expresión: para cada pico del espectro de magnitud de señal de corto tiempo, una distancia se refiere a la distancia de

expresión $vd(k_p)$ entre el espectro de señal en todo el pico y el espectro del ventaneo del Frame. Y se calcula así:

Ecuación 7

$$vd(k_p) = \left[\frac{1}{2L+1}\sum_{k=-L}^{L}\left(\left|s(k_p+k)\right| - |W(k)|\right)^2\right]^{\frac{1}{2}}$$

- Donde k_p es la frecuencia de índice de un pico espectral y L determina el número de componentes del espectro en cada longitud de todo pico que será comparado. El espectro de la señal $S(k)$, y el ventano de Frame $W(k)$, son normalizados para tener una magnitud igual a 1 en el pico que se uso en el cálculo de $vd(k_p)$. La distancia de expresión para los componentes de frecuencia en torno al pico representa las distancias de expresión del pico. Es decir:

Ecuación 8

$$vd(k) = vd(k_p)$$

Para todo

Ecuación 9

$$k \in [k_p - L, k_p + L]$$

- Note que si la información sobre la frecuencia fundamental es fácil de conseguir, la distancia de voz pudo ser calculada en los índices de frecuencia correspondientes a múltiplos de frecuencia fundamental en vez de los picos del espectro. También hay que notar que la estimación de la frecuencia fundamental pudo ser obtenida basándose sobre el mínimo acumulado de la distancia de expresión calculada en múltiplos de valores de frecuencia fundamental considerados (Garcia Luz, 2008).
- Cálculo de la distancia de expresión para canales de bancos de filtros: La expresión de distancia para canal FB (bancos de filtro) se calcula como media ponderada de las distancias de expresión dentro del canal, lo que refleja el cálculo de energías de los bancos de filtros que se utilizan para obtener las características de reconocimiento. Así (Garcia Luz, 2008):

Ecuación 10

$$vd^{fb}(b) = \frac{1}{X(b)}\sum_{k=k_b}^{k_b+N_b} vd(k)\cdot G_b(k)\cdot |S(k)|^2$$

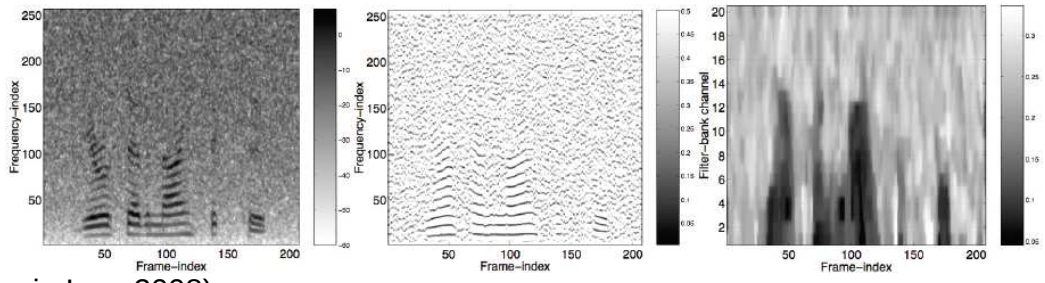
Donde

Ecuación 11

$$X(b) = \sum_{k=k_b}^{k_b+N_b-1} G_b(k) \cdot |S(k)|^2$$

Procesamiento posterior de la distancia de expresión: Expresando la distancia obtenida a partir de los pasos (2) y (3) puede accidentalmente ser de un valor bajo por una región sin expresión o viceversa. Para reducir estos errores, se debe haber filtrado las distancias de voz empleando filtros que calculan la mediana en 2-D debido a su eficacia en la eliminación de valores atípicos y simplicidad. En una configuración, el tamaño medio de los filtros de 5x9 y 3x3 (el primer número es el numero de Frames y el segundo es el numero de índices de frecuencia) se utilizan para filtrar la distancia de voz $vd(k)$ y $vd^{fb}(k)$, respectivamente. Ejemplos de espectrogramas de ruido de voz y la correspondiente distancia de voz por espectro y canales de banco de filtros representado en la figura 11 (Garcia Luz, 2008).

Figura 11. Representación de espectrogramas con algunas caracteristicas



(Garcia Luz, 2008)

La figura muestra las siguientes características. El espectrograma de la izquierda es el inicial en el análisis. El espectrograma del medio muestra la distancia de expresión en el dominio de frecuencia, y en el espectrograma de la derecha representa la voz dañada por el ruido blanco en SNR = 5 dB.

4.3.9 REVERBERACIÓN Y RUIDO

La comunicación de la voz es tan natural para todos los seres humanos y normalmente no suelen darse cuenta de algunos efectos. Antes de llegar a un micrófono o a los oídos del receptor, las señales del habla pueden modificarse por el medio en que se propaga (Lofqvist, 2006).

En una cámara anecoica (cámara libre de reflexiones y refracciones) ideal, la señal sólo sigue una ruta desde la fuente hasta el receptor. Sin embargo, en las habitaciones típicas, las superficies reflejan la emisión de sonido, el micrófono recibe una corriente reflejando las señales de múltiples caminos de propagación.

El conjunto de reflexiones que se denomina reverberación, la reverberación no es perjudicial en todo momento. Puede dar al oyente la impresión de espacio de caja sino que también aumenta el "liveness" y “warmth” especialmente en la música. Por otro lado, el exceso de reverberación es causa de pérdida de la inteligibilidad y la claridad, o perjudicar la comunicación musical.

El efecto de reverberación puede ser modelado como la transformación de una señal por un tiempo lineal invariante del sistema. Esta operación está representada por la convolución entre la sala de respuesta de impulso (RIR) y la señal original, expresado como (Lofqvist, 2006):

Ecuación 12

$$y(n) = x(n) * h(n)$$

Donde $y(n)$ representa la señal de voz degradada, $x(n)$ representa la señal de voz original y $h(n)$ denota la respuesta al impulso. Además de la reverberación, el sonido también está sujeto a la degradación por el ruido aditivo. Las fuentes de ruido, tales como ventiladores, motores, entre otros. Por lo tanto, se debe incluir el efecto del ruido sobre el modelo de señal degradada $y(n)$ siendo representada de la siguiente manera:

Ecuación 13

$$y(n) = x(n) * h(n) + v(n)$$

Donde $v(n)$ representa el ruido aditivo (Lofqvist, 2006).

4.3.10 SUSTRACCIÓN ESPECTRAL

La sustracción espectral es una técnica que permite la mejora de la voz, la cual hace parte de la clase de STSA (amplitud espectral de corto tiempo) lo que hace interesante la sustracción espectral es precisamente su simplicidad y baja complejidad computacional, además de su ventaja en la implementación en plataformas con recursos limitados. A continuación se dará una explicación acerca del algoritmo que se utiliza para lograr una correcta sustracción espectral.

Antes de presentar la substracción espectral como enfoque en cuanto a la dereverberación, se va revisar la formulación original como una técnica de reducción de ruido. La cual se puede expresar en el dominio de la frecuencia como (Garcia Luz, 2008):

Ecuación 14

$$Y(k) = X(k) + V(k)$$

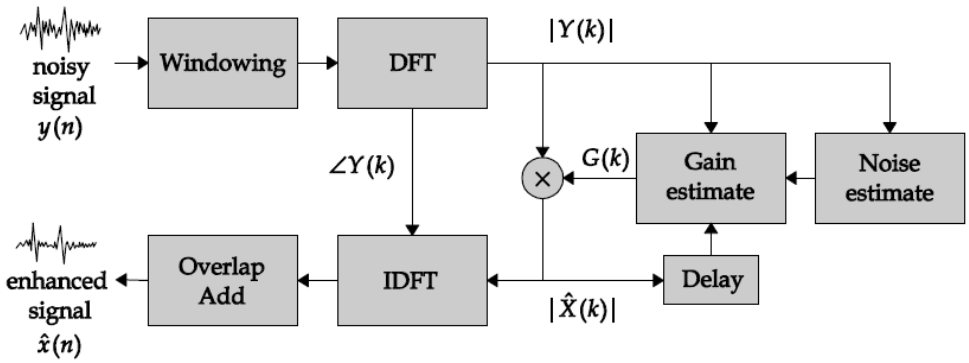
Esta ecuación denota la DTF (transformada discreta de Fourier en corto tiempo) de $y(n), x(n)$ y $v(n)$ respectivamente. La idea central de la sustracción espectral es recuperar $x(n)$ modificando sólo la magnitud de $Y(k)$. Este proceso puede ser descrito como una operación de filtrado espectral así (Garcia Luz, 2008):

Ecuación 15

$$\left|\hat{X}(k)\right|^{v} = G(k)|Y(k)|^{v}$$

Donde v denota el orden espectral, $\hat{X}(k)$ es la DFT de la señal mejorada $\hat{x}(n)$, y $G(k)$ es simplemente una ganancia de función.

Figura 12. Diagrama de Bloques de la Sustracción Espectral



(Garcia Luz, 2008)

La Figura 12 muestra el diagrama de bloques de un procedimiento general de la sustracción espectral. La señal de ruido $y(n)$ es ventaneada para posteriormente aplicarle la DFT. La función de ganancia $G(k)$ se calcula utilizando las muestras de magnitud de ruido, La señal previa de mayor magnitud y el ruido estadístico. Vale la pena aclarar que $\angle Y(k)$, esta fase se mantiene sin cambios siendo una entrada al bloque de la inversa de la DTF (IDTF) el aumento de señal se obtiene asociando la mayor magnitud y la fase de $Y(k)$, procesadas por el bloque de la IDFT a lo largo de una operación de superposición y adición esta ultima para compensar el ventaneo. Los bloques de ganancia y de ruido estimados son la parte más crítica del proceso y el éxito de esta técnica depende de una adecuada determinación de las ganancias (Garcia Luz, 2008).

Ahora para obtener la estimación más simple de la ganancia de $G(k)$ es necesario tener en cuenta las siguientes especificaciones:

Ecuación 16

$$G(k) = \begin{cases} 1 - \dfrac{1}{SNR(k)}, & SNR(k) > 1 \\ 0, & otherwise \end{cases}$$

Donde $SNR(k)$ es una medida de la señal a-posteriori de la relación señal-ruido y $\hat{V}(k)$ es el ruido estimado. Es necesario prevenir que $\left|\hat{X}(k)\right|$ sea negativo, debido a que se debe sujetar a las condiciones dadas para $G(k)$ se presentan algunos inconvenientes. Hay que notar que las ganancias se calculan para cada cuadro en cada índice de frecuencia de forma independiente. Observando la distribución de estas ganancias en una red de tiempo y frecuencia, se nota que en celdas vecinas pueden mostrar diferentes niveles de atenuación esta irregularidad en las ganancias dan lugar a tonos aleatorios en las frecuencias que aparecen y desaparecen rápidamente. Dando lugar a un molesto efecto llamado ruido musical. Estimaciones más elaboradas para $G(k)$ tienen como objetivo reducir el ruido musical. Un mejor enfoque para estimar la ganancia esta dado por (Garcia Luz, 2008):

Ecuación 17

$$G(k) = max\left\{\left|1 - \alpha\left(\frac{1}{SNR(k)}\right)^{\frac{v}{2}}\right|^{\frac{1}{v}}, \beta\right\}$$

Donde α y β son los factores de sobre-sustracción espectral. El factor de sobre-sustracción controla la reducción residual de ruido. Menores niveles de ruido son alcanzados con un mayor α , sin embargo, si α es muy grande la señal de voz se distorsiona. El factor espectral trabaja en la reducción del ruido musical, en una amplia banda de frecuencias. Una adecuada elección de β también es necesaria ya que se este valor es muy grande otros artefactos indeseados se hacen más evidentes. Es importante señalar que la distorsión de la voz y el ruido residual no pueden reducirse simultáneamente (Garcia Luz, 2008).

4.4 ESTADO DEL ARTE

Los sistemas de software QNX, el cual es líder mundial de sistemas operativos y middleware para telemática en el automóvil y mercado informativo, QNX® Aviage® anuncio que su paquete de procesamiento acústico ganó el premio Elektra 2008 por un sistema empotrado como producto del año. Organizado por Electronics Weekly. El premio Elektra reconoce la excelencia y la innovación electrónica (QNX, 2008).

El paquete de procesamiento acústico de QNX® Aviage® es un innovador producto que reduce drásticamente el costo y mejora la calidad de los sistemas de manos libres para el automóvil. El paquete utiliza algoritmos patentados para extraer la voz humana de los interiores de autos ruidosos, mejorando así la claridad de la persona en conversaciones telefónicas a través de manos libres (QNX, 2008).

El paquete reduce la necesidad de hardware dedicado y reduce el proceso de ajuste, haciendo del manos libres y el sistemas de reconocimiento de voz mas asequible para una amplia gama de automóviles, basado en tecnologías ya desplegadas en las plataformas de los vehículos Acura, Audi, BMW, Daimler, Fiat, GM, Honda, Hyundai, Porsche, y otros fabricantes de automóviles, el paquete dedicado reemplaza el hardware de procesamiento de voz con un pequeño y eficiente software de solución (QNX, 2008).

Altamente personalizable, el paquete ofrece una librería modular de software de algoritmos que los diseñadores pueden utilizar por separado o en combinación sobre las bases de los requisitos de las aplicaciones. Del tal forma algunas de las aplicaciones incluidas por el paquete se presentan en módulos de cancelación del eco acústico, supresión de ruido, control automático de ganancia, nivel de control dinámico, ecualización paramétrica, ampliación de ancho de banda y eliminación de buffet de viento. Esta arquitectura modular junto con el apoyo de múltiples procesadores y DSP’s, permite que los diseñadores actualicen, modifiquen y reutilicen el paquete a través de múltiples líneas de productos (QNX, 2008).

4.5 LIMITACIONES Y ALCANCES

Como fue expresado en los objetivos del proyecto, en los cuales el planteamiento de estos viene dado solo hasta el reconocimiento, es decir, si el patrón se encuentra o no en la señal muestreada, más no se encarga de la reconstrucción del fonema.

Debido a que la señal de voz es una señal pseudoaleatoria presenta muchos inconvenientes para su procesamiento, uno de ellos es tener en cuenta el locutor. Ya que el locutor es el que introduce la mayor variabilidad, puesto que este no siempre pronuncia de la misma forma las palabras generando un inconveniente y preparando a los sectores de la investigación en el procesamiento de señales para la realización de reconocedores de patrones de audio más robustos.

La cantidad de palabras dichas por el locutor incrementa la dificultad en la implementación del reconocedor de fonemas por dos motivos. El primero porque al aumentar el número de las palabras es más fácil que aparezcan palabras parecidas entre sí, y el segundo porque el tiempo de procesamiento se incrementa al aumentar el número de palabras.

Cuando en el reconocimiento se intenta en una pronunciación de palabras de forma continua esta se dificulta, debido a que el hablante no pronuncia la palabra de la misma forma, es decir, la palabra es representada de distinta forma si esta va al inicio, a la mitad o al final de la frese precisamente porque es demasiado complejo interpretar por medio del reconocedor el fonema predecesor o el fonema siguiente.

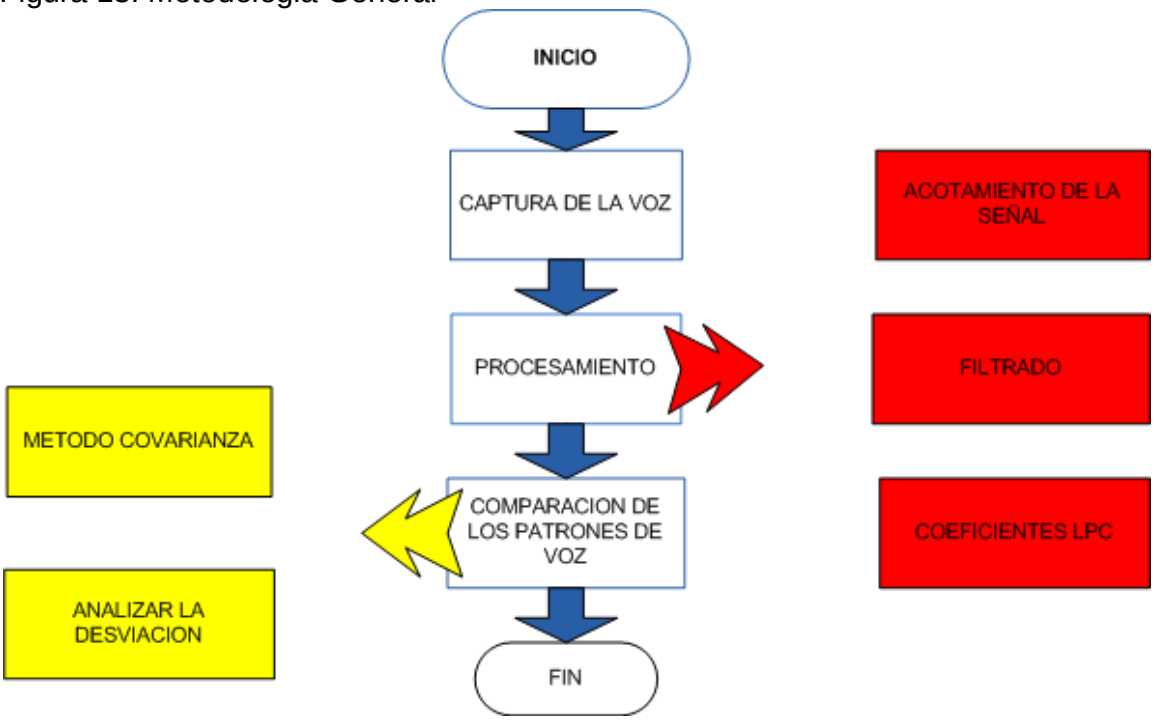
Para voces de frecuencia fundamental alta, la envolvente espectral aparece “muestreada” en pocos puntos, además no es posible separar la excitación de la envolvente espectral.

El reconocimiento automático de patrones de audio es un campo aún abierto a la investigación a nivel mundial debido a sus grandes complicaciones aleatorias que presenta. Actualmente existen diversos sistemas comerciales de reconocimiento de palabras continuas y discretas, implementados con diferentes métodos, pero estos presentan problemas y falta de robustez en situaciones reales, por lo que se sigue en la búsqueda de técnicas adecuadas en la selección de parámetros, que permita hacer óptima su compresión y clasificación.

5. DISEÑO METODOLÓGICO

La metodología que se llevó a cabo para el desarrollo del proyecto se muestra en la siguiente secuencia:

Figura 13. Metodologia General



Dentro de la secuencia general presentada en la figura 13 existen subdivisiones que representan los tres grandes procesos para que el desarrollo del proyecto fuese lo mas organizado posible.

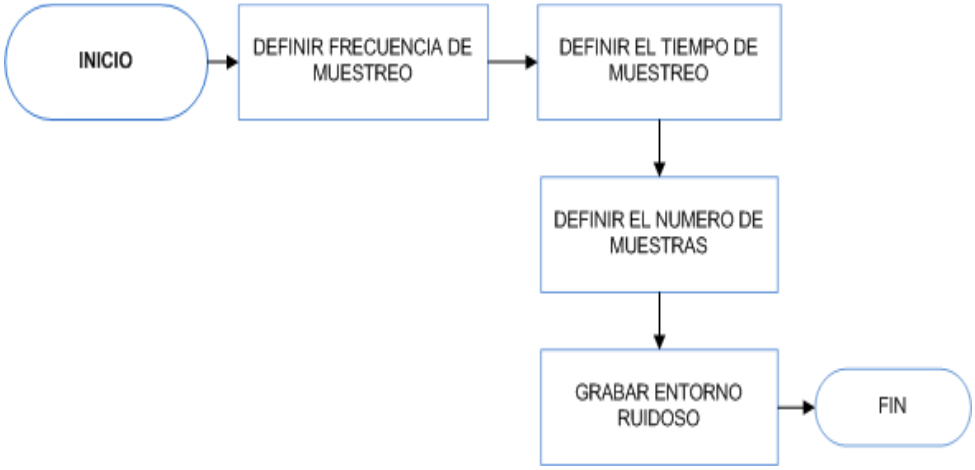
5.1 PASO 1 (CAPTURA DE VOZ (3 SEMANAS))

Es aquí donde se realiza la digitalización del archivo de audio para esto fue necesario definir la frecuencia de muestreo, el numero de muestras deseado y proceder con la grabación del archivo de audio.

- Frecuencia de muestreo: la selección de esta debe cumplir con el teorema de muestreo de Nyquist.
- Tiempo de muestreo: tiempo deseado de captura.
- Numero de muestras: se da por medio del producto de la frecuencia de muestreo con el tiempo de muestreo.

- Grabación del archivo de audio: representar la señal de audio en un archivo con extensión .wav.

Figura 14. Secuencia de Desarrollo Paso 1

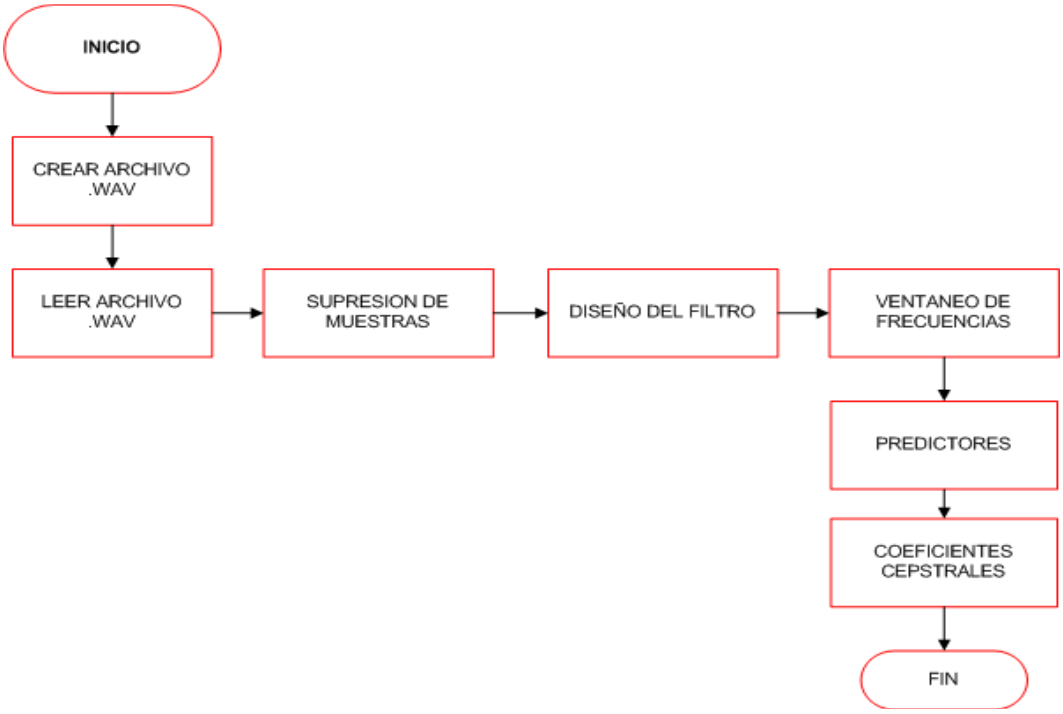


5.2 PASO 2 (PROCESAMIENTO DE LA SEÑAL (6 SEMANAS))

Luego de la digitalización de la señal de voz se procede con el procesamiento de la señal de audio el cual es representado así:

- Creación archivo .wav: el archivo no solo es representado como un gráfico, sino que la información que contiene este es representada en forma de vector.
- Lectura del archivo .wav: esto se realiza para manipular la información de la señal.
- Supresión de muestras: eliminación de cierta cantidad de muestras que no representan información útil.
- Diseño del filtro: creación e implementación de la herramienta de filtrado garantizando solo el rango de frecuencias de 200Hz a 8KHz.
- Ventaneo de frecuencias: división del vector resultante de la transformada discreta de Fourier de señal de audio para localizar las formantes de cada segmento.
- Predictores: generación de la función de transferencia a base de las formantes.
- Coeficientes cepstrales: se hallan para representar la envolvente espectral de la función de transferencia.

Figura 15. Secuencia de Desarrollo Paso 2

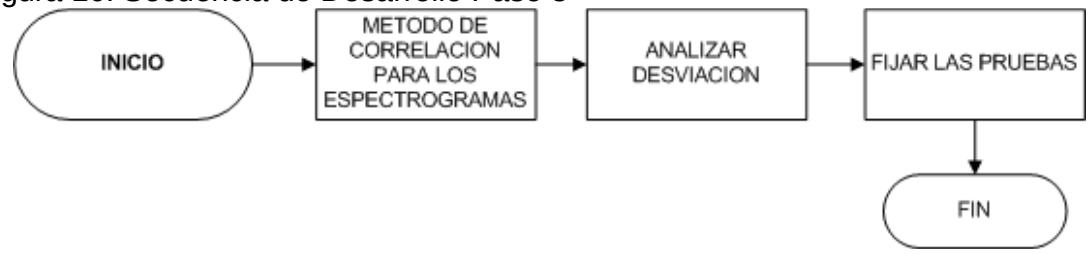


5.3 PASO 3 (COMPARACION DE LOS PATRONES DE VOZ (4 SEMANAS))

Luego de la representación de la envolvente espectral se procede con la comparación de los espectrogramas tanto de la señal de audio en entorno ruidoso como de la señal de audio en entorno sin ruido, esto con el fin de demostrar si el patrón se encuentra o no.

Para esto es necesario la implementación de ciertas técnicas matemáticas conocidas como correlación cruzada para determinar la desviación presentada en cada una de las pruebas determinando que si este valor es alto hay una alta probabilidad de que el patrón de voz si se encuentre en la señal muestreada.

Figura 16. Secuencia de Desarrollo Paso 3



5.4 PASO 4 (REALIAZACIÓN MONOGRAFIA (10 SEMANAS))

En el desarrollo de la monografía se invirtió un tiempo de 10 semanas abarcando el mayor tiempo del desarrollo total del proyecto, por ende, este fue realizado en paralelo con algunos de los pasos descritos en los numerales anteriores (5.1, 5.2, 5.3).

6. DESARROLLO

Un sistema de reconocimiento de patrones de audio tiene como principal función convertir la forma de onda de la voz en un algún tipo de representación paramétrica (obtener el espectros de frecuencias de la señal de audio), para así disminuir la cantidad de información de la señal y permitir su análisis y procesamiento. Posteriormente dichos espectros de frecuencia son confrontados con otros anteriormente obtenidos por medio de técnicas de reconocimiento.

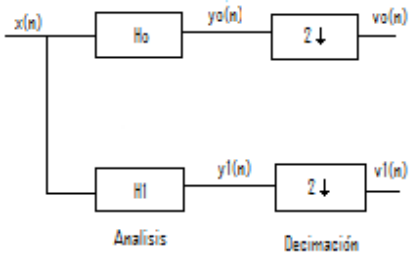
Existe gran variedad de representaciones paramétricas de una señal de voz entre las que se encuentran Bancos de filtros y modelo de Código de Predicción Lineal (LPC). Para la implementación de alguna de ellas hay que tener en cuenta los siguientes pasos:

- Medida del patrón de voz.
- Comparación con parámetros conocidos.

6.1 BANCO DE FILTROS

Los bancos de filtros se implementan para realizar descomposiciones de la señal original discretizada sobre un mismo segmento del espectro, es decir separar las frecuencias altas de las bajas, con el fin de buscar información dentro de las nuevas frecuencias obtenidas que permitan un reconocimiento del patrón auditivo como se muestra en la figura 17.

Figura 17. Banco de Análisis



Cuando se somete la señal a un banco de filtros se obtiene:

Ecuación 18

$$y_i(n) = x(n) * h_i(n)$$

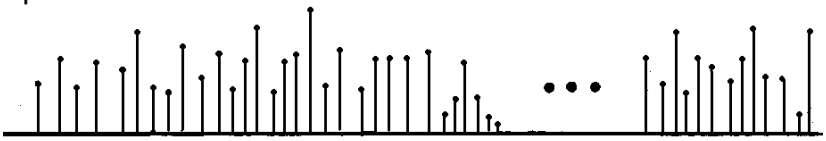
Donde

Ecuación 19

$$y_i(n) = \sum_{m=0}^{M_i-1} h_i(m)s(n-m)$$

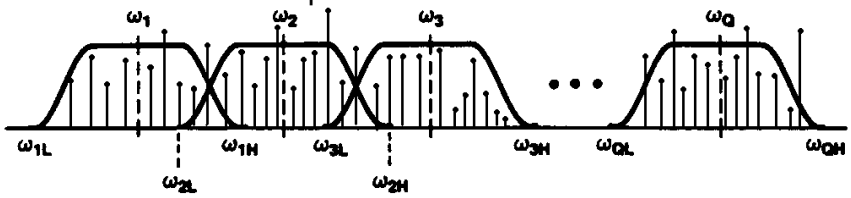
Las ecuaciones 18 y 19 anteriores significan la velocidad con la que se filtra o se muestrea la señal, involucran también el ancho de banda del filtro y el desplazamiento que este realiza sobre el espectro de la señal de voz, con el fin de separar las frecuencias altas de las bajas y obtener información en todos los segmentos de la señal. En la figura 18 se muestra el espectro de la señal original, cabe anotar que la suma de todos segmentos filtrados representan la totalidad del espectro de la voz.

Figura 18. Espectro de la Señal



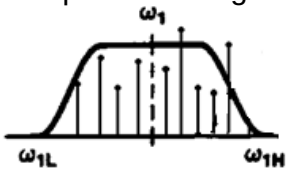
En la figura 19 se observa el muestreo realizado con los filtros sobre el espectro de la señal original.

Figura 19. Ventaneo Sobre el Espectro de Señal



En la figura 20 se muestra el resultado del ventaneo de cada filtro.

Figura 20. Resultado del Ventaneo por cada Segmento del Espectro de Señal



Cuando se filtra un segmento del espectro de la señal se genera una señal con un ancho de banda igual al del filtro (figura 20), posteriormente se efectúa un proceso

llamado decimacion, el cual depende del factor que lo acompañe que sirve para indicar cada cuanto debe suprimirse una muestra. Por ejemplo, Si se tiene un factor de decimación de 3 esto quiere decir que por cada 3 muestras se suprimirá una. Ahora bien, cuando se realiza el proceso de decimación, la señal filtrada se expande hasta el ancho de banda de la señal original (Uruguay, 2007).

6.2 LPC (CODIGO DE PREDICCIÓN LINEAL)

Dado que LPC es capaz de extraer la información lingüística y eliminar la correspondiente a la persona particular. La predicción lineal modela la zona vocal humana como una respuesta al impulso infinita, que produzca la señal de voz.

El término predicción lineal se refiere al método para predecir ó aproximar una muestra de una señal en el dominio del tiempo s[n] basada en varias muestras anteriores s[n - 1], s[n - 2], s[n - M].

Ecuación 20

$$s[n] \approx \hat{s}[n] = -\sum_{i=1}^{M} a_i s[n-1]$$

Donde s[n] es llamada señal muestreada, y a_i, i = 1,2,…, M son los predictores ó coeficientes LPC. Un pequeño número de coeficientes LPC $a_1, a_2, \ldots, a_M$ pueden ser usados para representar eficientemente una señal s[n]. Los valores a_1 $a_2, \ldots, a_M$ son la base para la realización de este trabajo debido a que nos ayudan a modelar los parámetros de la voz de cada uno de los hablantes que se emplean en este sistema propuesto (Lopez, 2005).

6.2.1 PREDICCIÓN LINEAL DE LA PARTE CAUSAL DE LA AUTOCORRELACION

A partir de la secuencia de autocorrelación $R(n)$ definimos su parte causal como:

Ecuación 21

$$R^{+}(n) = \begin{cases} R(n); & n > 0 \\ \dfrac{R(0)}{2}; & n = 0 \\ 0; & n < 0 \end{cases}$$

Su transformada de Fourier en el espectro complejo.

Ecuación 22

$$S^{+}(\omega) = \frac{1}{2}[S(\omega) + S_H(\omega)]$$

Donde $S(\omega)$ es el espectro, es decir, la transformada de Fourier de $R(n)$**,** y $S_H(\omega)$ es la transformada Hilbert de $S(\omega)$.

Debido a la analogía entre $S^{+}(\omega)$ y la señal analítica usada en modulación de amplitud, se puede definir una envolvente espectral como.

Ecuación 23

$$E(\omega) = |S^{+}(\omega)|$$

La envolvente espectral de la señal se estima por un filtro digital IIR todo polo. En codificación lineal predictiva, el sistema todo-polos reemplaza el banco de filtros paso-banda de su predecesor y se usa en el *encoder* para blanquear la señal (aplanar su espectro) y de nuevo en el decodificador para reasignar la envolvente espectral de la señal de voz original.

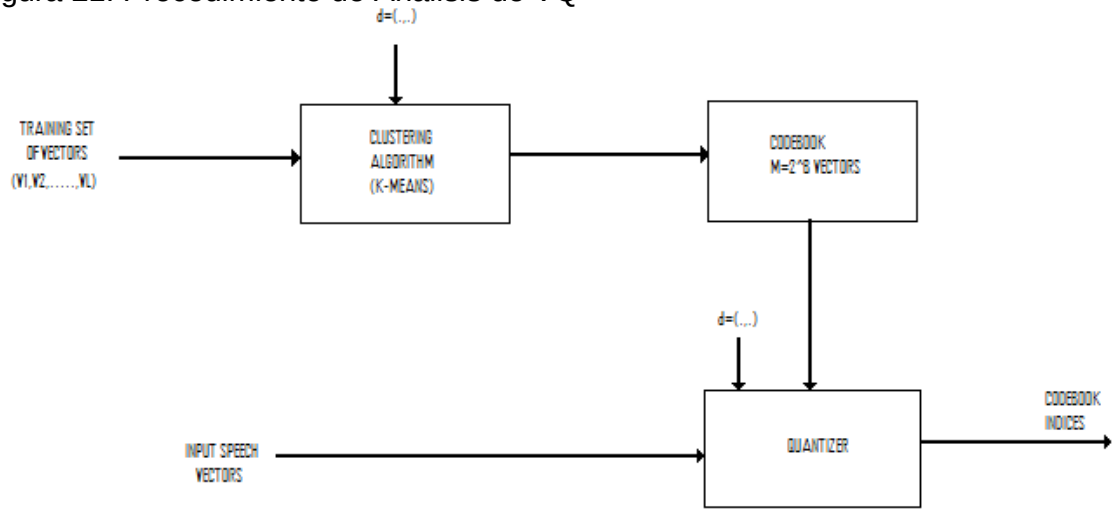
Esta característica de envolvente, junto al alto rango dinámico del espectro de la señal de voz, origina que el cuadrado de la envolvente espectral $E^2(\omega)$, que es además el espectro $R^{+}(n)$, sea más robusto al ruido que el propio espectro. Además, es un hecho bien conocido que $R^{+}(n)$ tiene los mismos polos y la misma multiplicidad que la señal.

Ambas propiedades conducen a considerar que la predicción lineal de $R^{+}(n)$ como una técnica robusta de representación de la señal de voz. Al igual que la técnica LPC estándar asume un modelo todo polo para $S(\omega)$, esta nueva técnica equivale un sistema todo polo para $E^2(\omega)$. Ello da lugar a que esta técnica solo realice una deconvolución parcial de la señal de voz.

6.3 VQ (VECTOR DE CUANTIZACION)

Para la construcción del codebook de un vector de cuantización se debe tener en cuenta la representación de la figura 21:

Figura 21. Procedimiento de Análisis de VQ



(Rabiner, 1993)

Un gran conjunto de vectores de prueba denotados como $v_1, v_2, v_3, \ldots \ldots v_L$ son usados para crear un buen codebook de vectores para la representación de la variabilidad espectral. Si la longitud del codebook del VQ está dada por $M = 2^B$ vectores, entonces se requiere que $L \gg M$ para poder encontrar un buen conjunto de M vectores del codebook. En la práctica suele utilizarse un valor de $L = 10M$ para que el codebook del VQ se comporte razonablemente bien (Rabiner, 1993).

Una medida de similitud o de distancia entre un par de vectores de análisis espectral podría ser representada por un conjunto de vectores prueba Cluster el cual asocia o clasifica vectores espectrales arbitrarios al interior de un único codebook de entradas. Para esto se debe saber que la distancia entre dos vectores se da por medio de:

Ecuación 24

$$d(v_1, v_2) = \begin{cases} 0 \; si \; v_1 = v_2 \\ > 0 \quad EOC \end{cases}$$

Los cuantizadores escalares se basan en que un valor de salida es resultante de la muestra de entrada en ese momento y, quizá de N muestras de entrada previas. Los cuantizadores por bloques o vectoriales toman N_1 muestras de entrada a la vez y estas son mapeadas a un vector de dimensión N_2, donde $N_2 \leq N_1$. De esta forma se crea un vector de cuantización o cantroide.

Ahora bien, se debe calcular un vector de cuantización tanto para la señal de voz capturada en un entorno ruidoso como para la señal capturada en un entorno sin

ruido los cuales serán comparados para determinar si el patrón de voz se encuentra en la señal de entorno ruidoso o no.

6.4 CAPTACION DE LA VOZ

Inicialmente el proceso de captura de voz se realizó por medio de la herramienta de grabación de voz ofrecido por Windows generando como resultado un archivo de audio con extensión .WMA. Al momento de Importar el archivo a Matlab el software no permitió leer la información contenida en el archivo con esta extensión, lo cual impedía continuar con el desarrollo de la captura de voz.

6.4.1 FORMATO WMA (WINDOWS MEDIA AUDIO)

Windows Media Audio o **WMA** es un formato de compresión de audio con pérdida aunque se ha avanzado en la compresión de audio con **LOSELESS (Menos pérdida)**. El códec ofrece una gama dinámica de control utilizando el máximo y el promedio de las amplitudes de audio que se calculan durante el proceso de codificación para evitar así menos perdida en la compresión (Media, 2008).

Una tasa baja de bits codificada está basada en el modo mixto que incluye voz y música; el códec aprovecha el rango de frecuencia de la voz humana con el fin de maximizar la compresión y en el modo música funciona como un códec estándar de Windows Media Audio. La codificación del contenido está configurada de tal forma tal forma que se pueda cambiar automáticamente entre modos (Media, 2008).

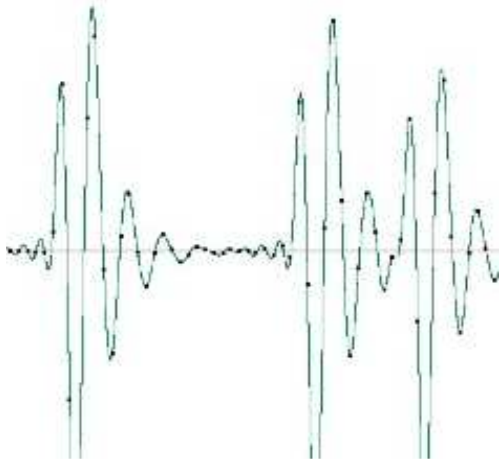
Debido a estas características este formato de audio es uno de los más apropiados para la implementación de este proyecto, pero como Matlab no reconoce archivos con extensión .WMA fue necesario buscar un formato de audio que aceptara este software (Media, 2008).

Matlab ofrece dentro del toolbox de procesamiento digital de señales una serie de instrucciones para capturar señales de audio creando un archivo con extensión .WAV.

6.4.2 FORMATO WAV (WAVEFORM AUDIO FORMAT)

El nombre que se le ha dado es (Wave audio format) o (formato de audio en forma de ondas). Una de las características principales de los sonidos WAV es que no se encuentran comprimidos por lo cual ofrecen la reproducción de los sonidos originales en la máxima calidad posible (Kioskea, 2007).

Figura 22. Tipica manifestacion gráfica de un sonido



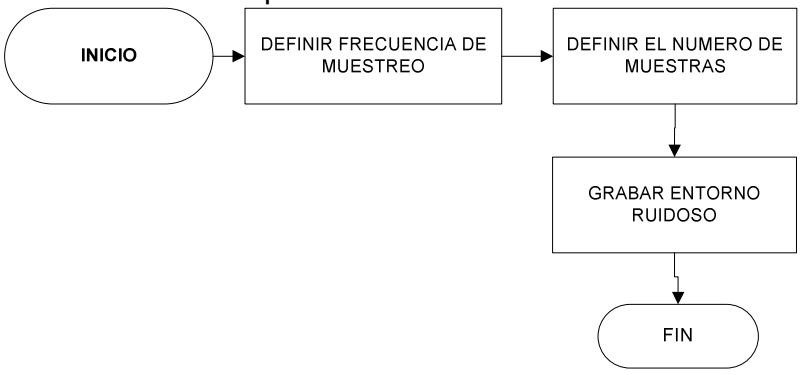
(Pianored, 2004)

Debido a que este tipo de archivo no se encuentra comprimido no son muy populares en Internet debido al gran tamaño que poseen por tanto recolecta mas muestras y almacenando mas información (Kioskea, 2007).

A pesar de que el formato WAV puede soportar casi cualquier códec de audio, se utiliza principalmente con el formato PCM (modulación por codificación de pulso no comprimido). Por cada minuto de grabación de sonido se consumen unos 10 megabytes de espacio en disco. Una de sus grandes limitaciones es que solo se puede grabar un archivo de hasta 4 gigabytes, que equivale aproximadamente a 6,6 horas en calidad de CD de audio (Kioskea, 2007).

6.4.3 CAPTACIÓN DE LA VOZ EN MATLAB

Figura 23. Secuencia captura de voz



Dentro del proceso de captura de voz en Matlab como primera medida hay que definir la frecuencia de muestreo Fs, para esto es necesario conocer unos rangos de frecuencia o anchos de banda como:

- Ancho de banda de la voz.
- Ancho de banda del oído humano.
- Espectro electromagnético del audio.

Para el desarrollo del proyecto requirió dimensionar que rango de frecuencias no fueron tenidas en cuenta. De esta forma se debió conocer cuáles son los respectivos anchos de banda, vale la pena aclarar que estos rangos se encuentran dentro de VLF (Very Low Frecuency).

Espectro electromagnético del audio: está compuesto por frecuencias que varían desde los 3 Hz y los 30 KHz.

Ancho de banda del oído humano: de una forma ideal se dice que el oído humano percibe frecuencias entre 20 Hz y los 20 KHz.

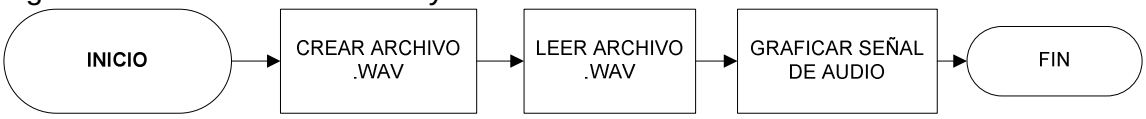
Ancho de banda de la voz: este rango de frecuencias se encuentra entre los 200 Hz y los 4 KHz.

Ya cuando se obtuvo esta información se definió una frecuencia de muestreo Fs=20000 para cumplir con el ancho de banda del oído así sea de una forma ideal. Posteriormente a la selección de la frecuencia de muestreo fue necesario definir un tiempo de captura, el cual depende de la duración del archivo de audio que deseemos obtener.

Cuando se habla de la definición del número de muestras simplemente se realizó como el producto entre la frecuencia de muestreo con el tiempo de captura definido. Y finalmente con el comando wavrecord en Matlab proceder a la grabación del archivo de voz.

6.4.4 CREACIÓN, LECTURA Y GRAFICA ARCHIVO .WAV

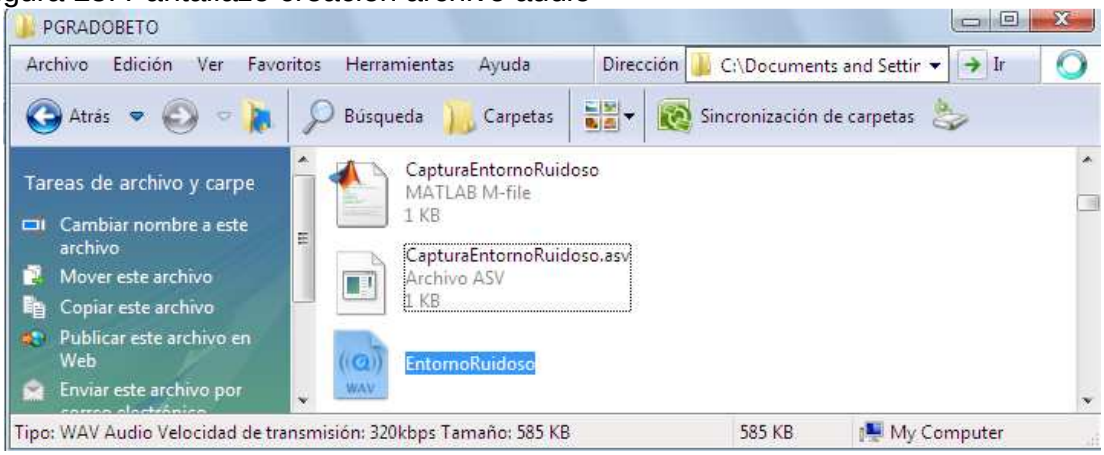
Figura 24. Secuencia creación y lectura del archivo de voz



En cuanto a la creación y lectura de un archivo .WAV simplemente fue posible gracias a la implementación de dos comandos que ofrece el toolbox de procesamiento digital de señales.

Para la creación de archivo .WAV se planteó el comando wavwrite el cual junto con los datos recolectados con wavrecord y la frecuencia de muestreo genera un archivo con el nombre que se deseo, con extensión .WAV. Para este caso el archivo generado es EntornoRuidoso.wav.

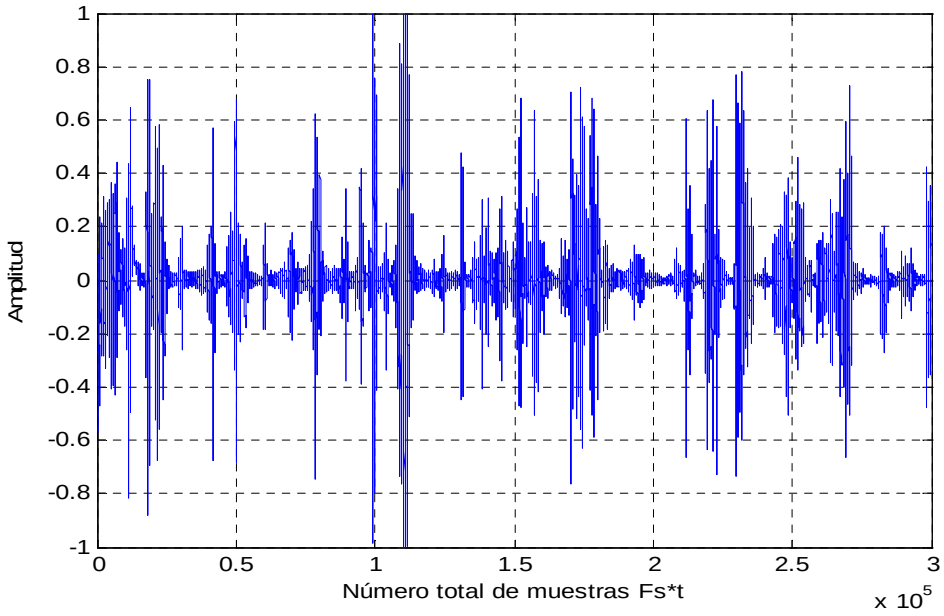
Figura 25. Pantallazo creación archivo audio



Luego de haber creado el archivo, se procedió con la lectura del mismo, esto se realizó con el comando wavread el cual guarda cada muestra de la señal en una posición de un vector.

Para graficar la señal se tiene que por defecto Matlab asignará valores máximo entre 1 y -1 para las amplitudes de cada una de las muestras (Figura 26). Con esto el trabajo restante para graficar la información del archivo de audio fue asignar el valor de la muestra con respecto al número total de las mismas.

Figura 26. Señal de voz



6.5 FILTRADO DE LA VOZ

Debido a que en el proyecto la captura de voz se realizó en un entorno donde no solo existía esta señal, esto produce que ruido aditivo generado por todas las perturbaciones alrededor del transductor o micrófono sean inherentes y afecten las señales de voz y el reconocimiento de las mismas.

Es por eso que inicialmente para solucionar este problema se diseño un filtro cuyas frecuencias de corte fuesen lo más similares a las del ancho de banda de la voz (200Hz – 4KHz). Con el fin de suprimir todas aquellas frecuencias que no se encontraran dentro de este rango y que pudiesen afectar posteriormente el reconocimiento de los patrones de voz.

6.5.1 FILTROS DIGITALES

Un filtro es básicamente una caja negra con una entrada y una salida. Si la salida es diferente a la entrada, significa que la señal original ha sido filtrada. Cualquier medio por el cual una señal de audio pasa, cualquiera sea su forma, puede describirse como un filtro.

Un filtro digital es un sistema de tiempo discreto que deja pasar ciertos componentes de frecuencia de una secuencia de entrada sin distorsión y bloquea

o atenúa otros. Se trata simplemente de un filtro que opera sobre señales digitales, tales como las que operan dentro de un computador. Un filtro digital es una computación que recibe una secuencia de números (la señal de entrada) y produce una nueva (la señal de salida) (Cadiz, 2003).

El filtro digital ofrece una serie de ventajas las cuales son:

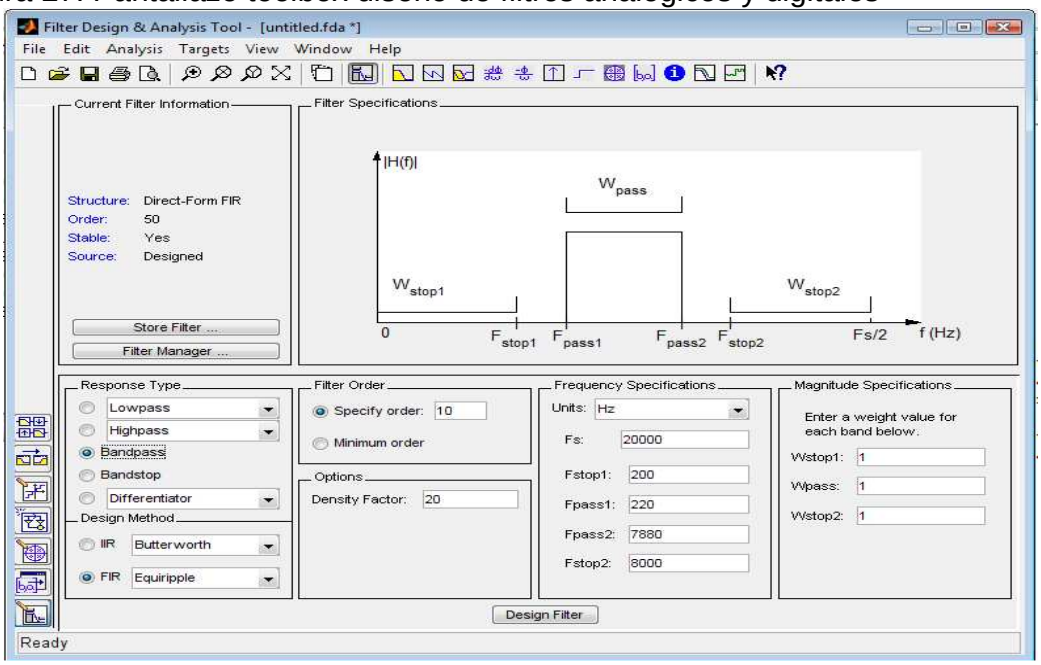
- Un filtro digital es ***programable***, su función está determinada por un programa almacenado en el procesador. Esto significa que el efecto del filtro puede ser cambiado fácilmente sin modificar su circuitería (Slideshare, 2005).
- En cambio, los filtros digitales no sufren problemas de variación de temperatura o de velocidad, y son extremadamente ***estables*** con respecto al tiempo (Slideshare, 2005).

Los filtros digitales son mucho más ***versátiles*** en su capacidad de procesar señales de diferentes formas. Esto significa que algunos filtros digitales tienen la capacidad de adaptarse a los cambios en las características de la señal (Slideshare, 2005).

6.5.2 DISEÑO DEL FILTRO

Gracias a que Matlab ofrece diversas herramientas de diseño para múltiples necesidades se recurrió al uso del toolbox de diseño de filtros. Esta herramienta tiene como nombre Filter Desing & Analysis Tool. La ventana que ofrece la herramienta de diseño del filtro se observa en la figura 27:

Figura 27. Pantallazo toolbox diseño de filtros analógicos y digitales



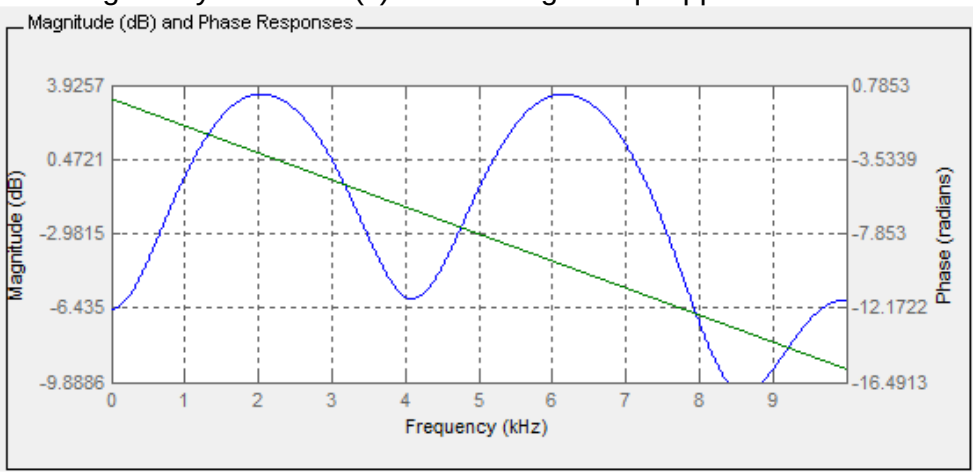
Para diseñar el filtro se definen las frecuencias de corte, se especifica el orden, si se quiere pasa banda o rechaza banda, si se quiere el filtro analógico o digital, además se puede observar que esta herramienta permite mostrar la respuesta en magnitud, la respuesta en fase, diagrama de polos y ceros, la respuesta al pulso y la respuesta al paso.

El filtro que se diseño para suprimir las frecuencias fuera del ancho de banda de la voz tiene los siguientes parámetros:

- Filtro digital FIR Equiripple.
- Pasa banda.
- Orden 10.
- Factor de densidad 20.
- Frecuencia Stop1 200Hz.
- Frecuencia Pass1 220Hz.
- Frecuencia Pass2 7880Hz
- Frecuencia Stop2 8000Hz
- Frecuencia de muestreo 20000Hz

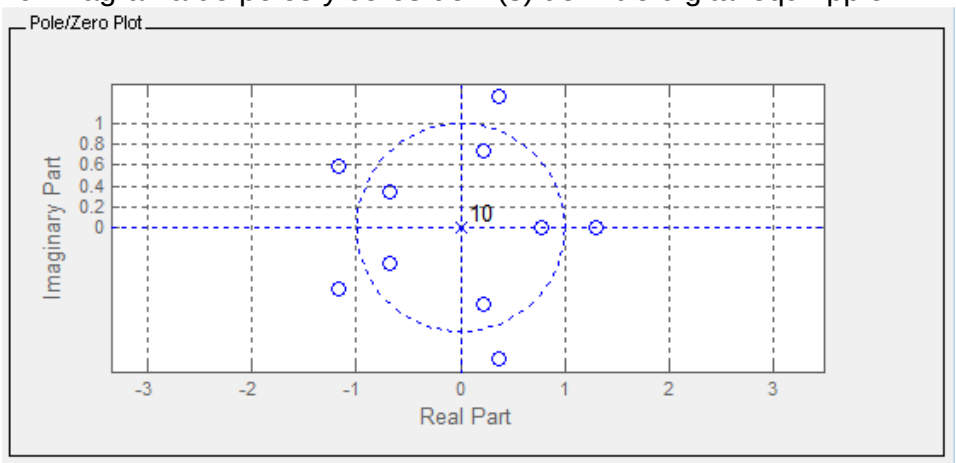
Con estas características se obtuvieron las siguientes respuestas respecto a Magnitud Y Fase (figura 28):

Figura 28. Magnitud y fase de H(s) del filtro digital equiripple



El filtro digital Equiripple es denominado así como debido a su respuesta en magnitud la cual presenta un rizado (RIPPLE) tanto en la banda de paso como en la banda de rechazo. Todos los rizados en la banda de paso tienen exactamente las mismas características es decir, tienen la misma magnitud y ocupan el mismo ancho de banda en rangos distintos de frecuencia. Al igual que en la banda de paso la banda de rechazo se obtiene las mismas características. Debido a este a este tipo de respuesta fue atribuido su nombre Equiripple (RIZADO EQUITATIVO).

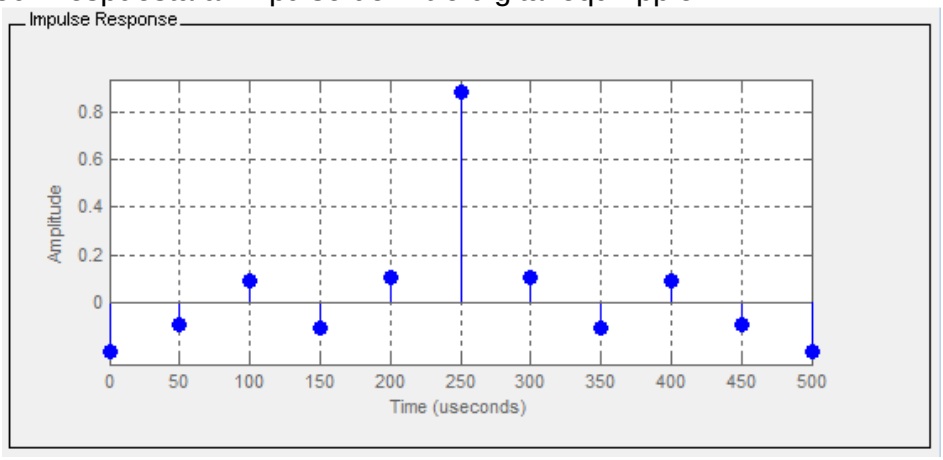
Figura 29. Diagrama de polos y ceros de H(s) del filtro digital equiripple



Cuando se analiza un el diagrama de polos y ceros de un filtro digital es para determinar si el sistema presenta inestabilidad de algún tipo. Este análisis es posible de realizar cuando al graficar los polos y los ceros de la función de transferencia del filtro estos no se ubican sobre un círculo de radio=1. En este

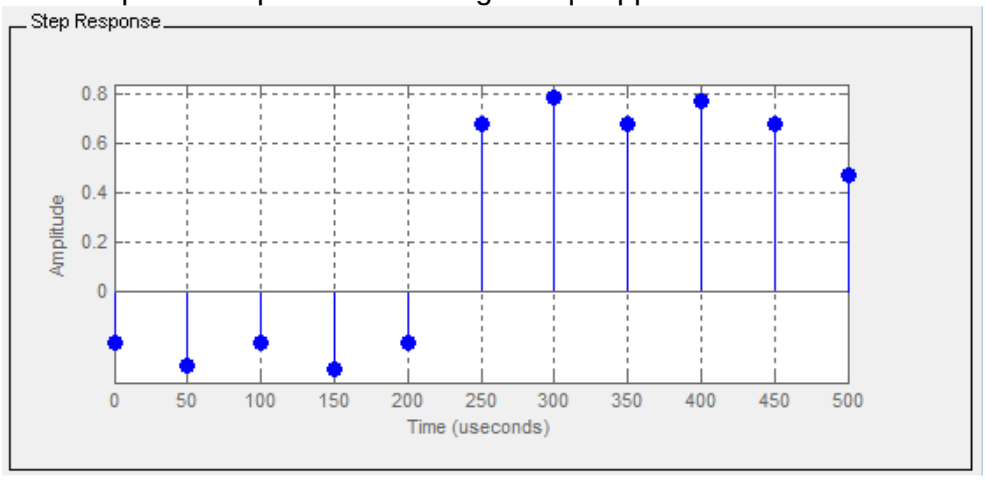
caso el sistema es estable, por el contrario si algún polo o cero se sobrepone en el círculo el sistema será inestable.

Figura 30. Respuesta al impulso del filtro digital equiripple



La respuesta al impulso está dada por los mismos parámetros de diseño del filtro, ya que la técnica consiste en derivar primero la función de transferencia del filtro, la cual es periódica en la frecuencia de muestreo. Así la función de transferencia puede ser expresada en series de Fourier. De esta forma se observa que los coeficientes de Fourier resultantes de la expansión dan como resultado la respuesta al impulso del filtro.

Figura 31. Respuesta al paso del filtro digital equiripple



Recordando que las respuestas al paso e impulso contienen información idéntica, pero en diferentes formatos. La respuesta al paso es útil en los análisis en el

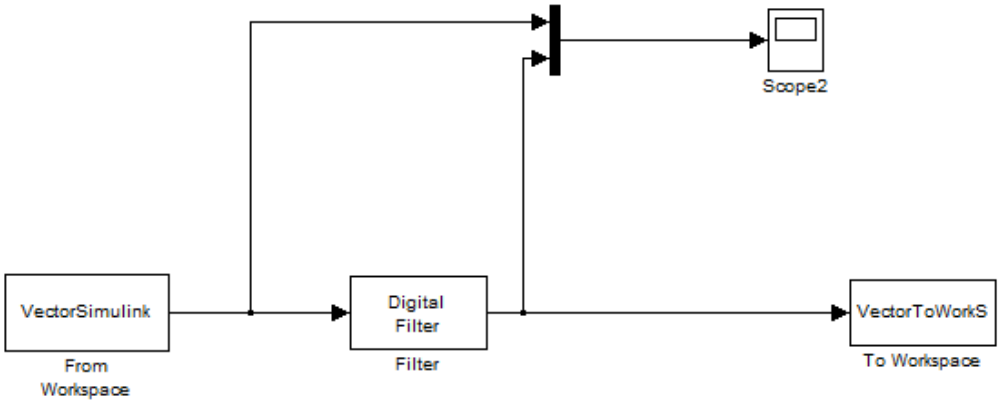
dominio del tiempo pues coincide con la forma humana de observar la información proporcionada por una señal.

Los parámetros de la respuesta al paso son importantes en el diseño de un filtro para la duración de la respuesta, ya que la respuesta al paso debe ser lo más rápida posible. Lo que es notorio en la respuesta al paso del filtro diseñado.

6.5.3 CREACIÓN DEL BLOQUE DE FILTRADO

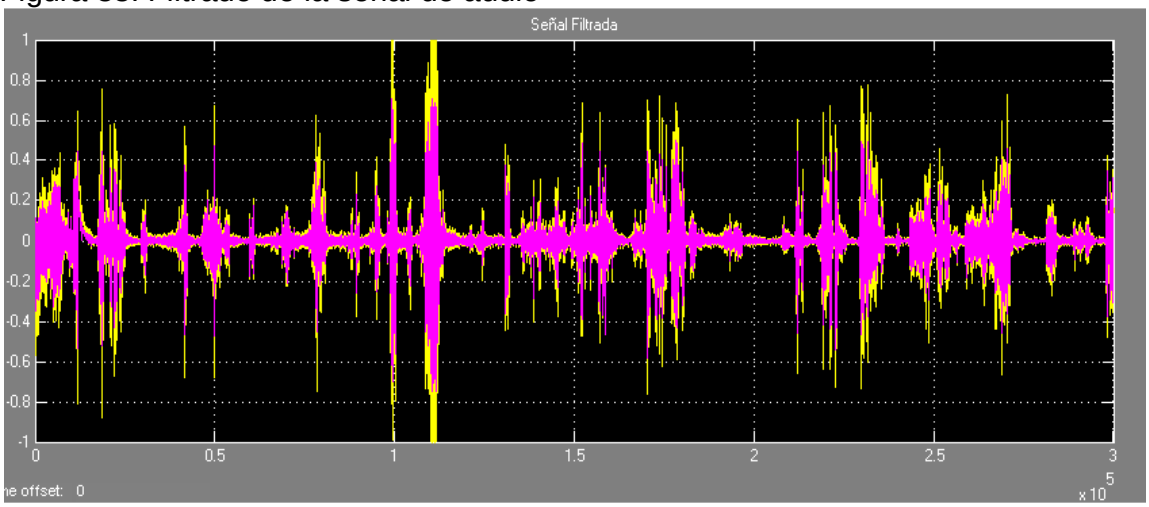
La herramienta de diseño del filtro nos permite exportar el filtro como un bloque a una subdivisión de Matlab llamada Simulink. Pero para poder filtrar el archivo de audio que contiene la información del entorno ruidoso este también debe ser exportado a Simulink como se muestra en la figura 32.

Figura 32. Diagrama de bloques del filtro



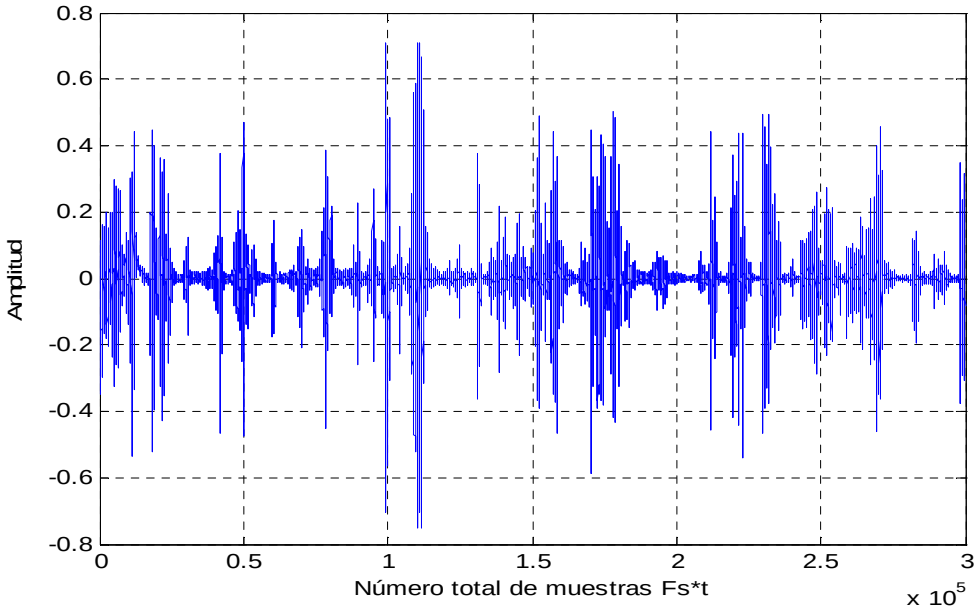
El resultado del filtrado se verá en la figura 33 la cual representa las muestras de audio por medio del color amarillo. Ahora por medio del color morado se ve la señal filtrada.

Figura 33. Filtrado de la señal de audio



De tal forma la señal filtrada será representada así (figura 34):

Figura 34. Señal filtrada



Luego del respectivo filtrado de la señal capturada se procedió a un análisis sinusoidal el cual está compuesto de una serie de modelos matemáticos los cuales representan la implementación del algoritmo para el reconocimiento de patrones de voz.

En los sistemas de reconocimiento de voz no se intenta, como mucha gente piensa, reconocer los sonidos del fonema, sino identificar una serie características principales para saber si la persona que habla dijo lo que se cree.

Hay que tener en cuenta que para la implementación del algoritmo existen factores que pueden implicar que el reconocimiento de los patrones de audio sea más complejo. A continuación de mencionarán algunos de estos factores:

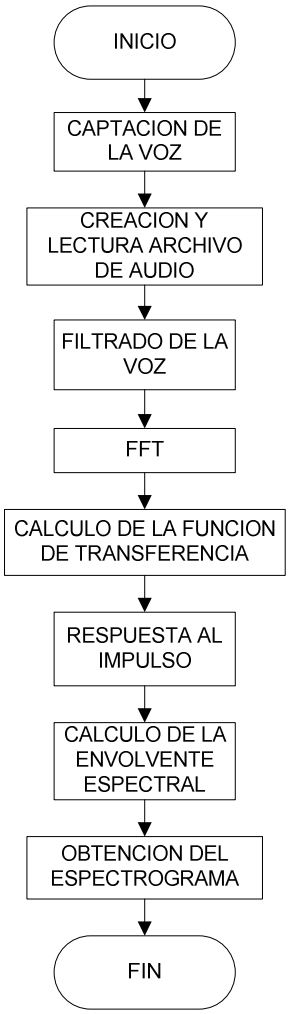
- Tamaño de la frase: implica que entre más larga sea la frase más difícil es el reconocimiento.
- Locutor: esto debido a que uno no pronuncia las palabras siempre de la misma forma. Esto incluye si la palabra va al inicio, en medio, o al final de la oración.
- Entorno físico: esto se debe al hecho de que no es lo mismo un sistema que funciona en un ambiente poco ruidoso, o por el contrario en un ambiente ruidoso.

Ahora bien luego de haber identificado algunos de los factores a tener en cuenta luego de la captura de voz se requirió la implementación de un código de predicción lineal (LPC), ya que una gran parte del tratamiento de la voz depende de este.

6.6 ANALISIS SINUSOIDAL DE LA VOZ

Debido a que la señal de voz no posee frecuencias sinusoidales fijas, esto las hace no estacionaria y no lineal. Para esto existen métodos que permiten representar las señales no estacionarias y no lineales como la suma de componentes que intervienes en la señal, generalmente segmentos de señal que no están afectados o que no presentan algún tipo de degeneración. Para esto se implemento el código de predicción lineal, el cual es descrito en la figura 35.

Figura 35. Secuencia Análisis LPC



Inicialmente la señal de voz en entorno ruidoso y en entorno sin ruido fue muestreada pasándola del dominio del tiempo continuo al dominio del tiempo discreto, cumpliendo con la teoría del muestreo de Nyquist, la cual dice que para que una señal sea muestreada con una buena cantidad de información la frecuencia de muestreo debe ser por lo menos del doble en comparación de la señal en el tiempo continuo, por ejemplo el ancho de banda de la voz es de 4KHz por tanto la frecuencia de muestreo debe ser de mínimo 8KHz. Esto se hace para que el cálculo de DTF (Transformada Discreta de Fourier) tenga un resultado coherente.

Ahora bien, prosiguiendo con el desarrollo se creó el archivo de audio tanto para la muestra de voz en un entorno ruidoso, como en un entorno sin ruido esto para poder realizarle los procesamientos necesarios dentro de los cuales se encuentran la lectura de los valores que representan la amplitud por cada instante de tiempo,

el filtrado de la voz para garantizar el trabajo solo con las características de la señal que se encuentren entre los 200Hz y los 4Hz y poder suprimir algunas degradaciones que se presenten en la señal muestreada que se encuentren por fuera del ancho de banda, claro que este filtrado no garantiza que la señal resultante ya esté libre de degradaciones ya que se pueden presentar perturbaciones de baja frecuencia al interior del rango de frecuencias que maneja la voz.

El algoritmo de predicción lineal parte del cálculo de la función de trasferencia del cuantizador de la voz (contador de muestras de la señal de voz), con el fin de poder calcular la respuesta la impulso para determinar la envolvente del espectro y por último a partir de estos, graficar el espectro de voz (Dominguez, 2001).

La transformada discreta de Fourier es una herramienta indispensable en los algoritmos LPC, debido a que se puede muestrear la señal y cuantizar los coeficientes digitales, ya que la implementación de esta es con bancos de filtros de análisis que descomponen la señal en ventanas ajustadas al ancho de banda de los filtros. Es aquí donde se determina el comportamiento de la envolvente del espectro que se generará, ya que se puede determinar la posición en frecuencia de las componentes formantes de la voz (estas son las que tienen mayor valor de amplitud) y son las que rigen el comportamiento espectral de la voz. A continuación se puede observar en la figura 36 la señal de voz en el dominio del tiempo y posteriormente en la figura 37 su respectiva representación de magnitud frente a frecuencia luego de que se le aplica la DTF.

Figura 36. Señal en el dominio del tiempo

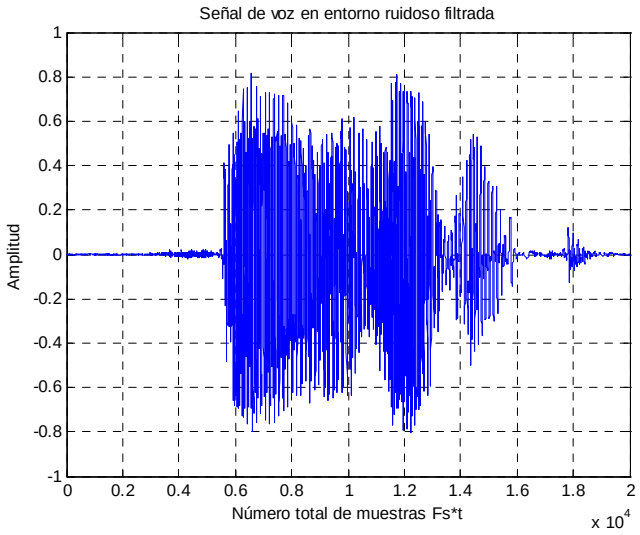
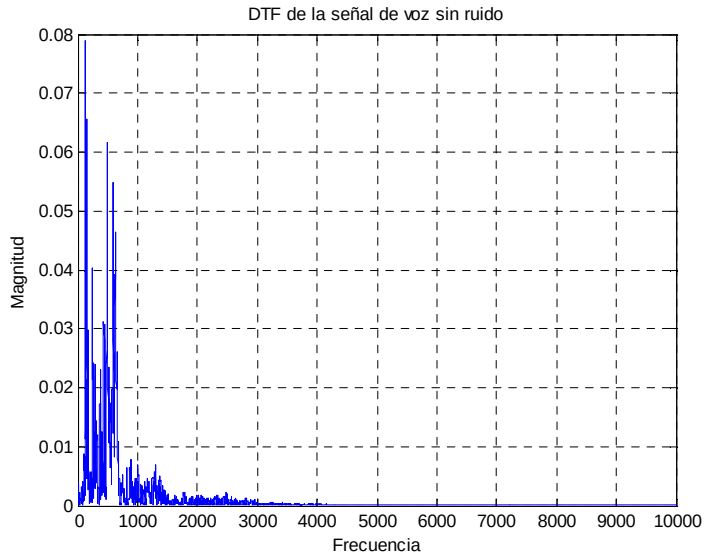
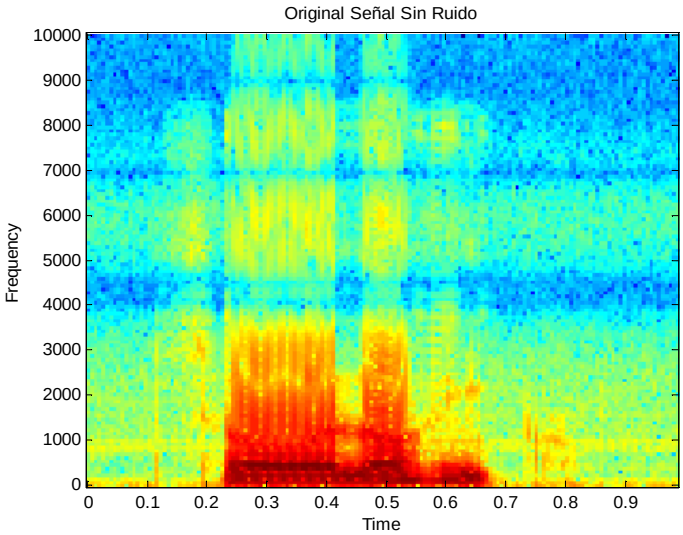


Figura 37. Señal en el dominio de la frecuencia



El espectrograma de la voz nos indica que en ciertas porciones de la señal se encuentra la mayor concentración de energía, es decir, los espacios en que las formantes son más grandes obteniendo valores pronunciados de magnitud teniendo la mayor energía se denota como una mancha roja en la gráfica del espectrograma, por el contrario en el lugar donde se representa la menor cantidad de energía se representa con un color azul como se ve en la figura 38 (Sancho, 2007).

Figura 38. Espectrograma de la señal de audio



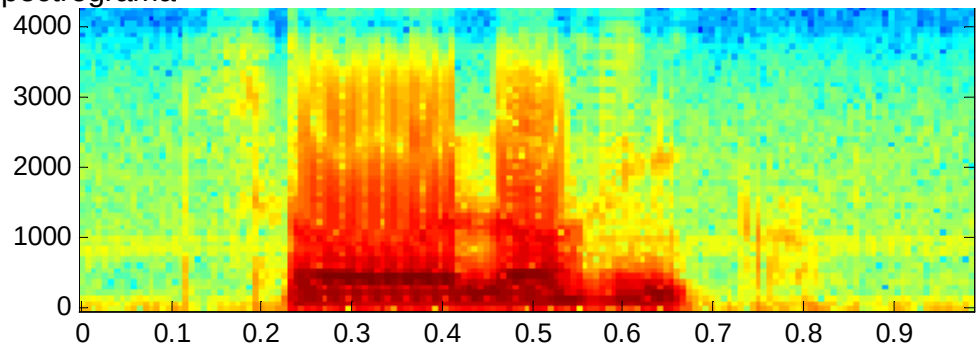
Vale la pena resaltar que un espectrograma de frecuencias es el resultado de calcular el espectro en tramas de ventaneo de una señal, esta se representa en una grafica tridimensional en donde se muestra la energía del contenido en frecuencias de la señal según va variando esta a lo largo del tiempo.

Dentro de la obtención del espectrograma Si se aplica una ventana muy grande obtendremos un espectrograma muy detallado pero a costa de incrementar el tiempo de cálculo necesario para esta operación. Para el caso de una ventana demasiado pequeña el efecto es el inverso y no seremos capaces de distinguir los diferentes armónicos si están muy juntos en el espectrograma.

El espectrograma sirve para analizar la sonoridad, la duración, la estructura de los formantes (timbre), la intensidad, las pausas, y el ritmo.

Puede observarse que en el espectrograma mostrado en la figura 39 existen dos bandas de frecuencia fundamentales que es donde se presenta la mayor cantidad de energía. Esta banda de frecuencia se encuentra entre los 0Hz y los 4000Hz.

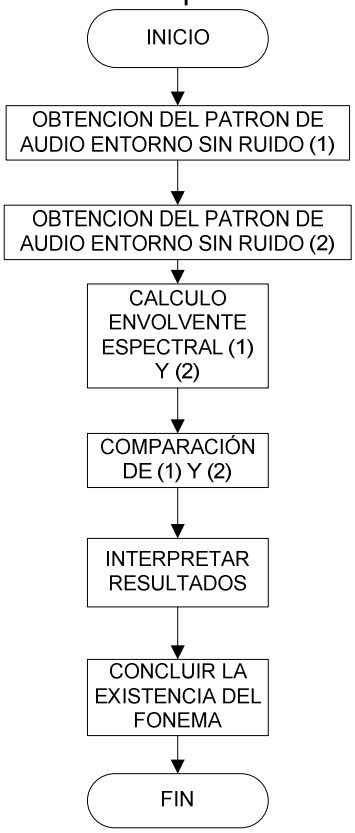
Figura 39. Bandas de frecuencia con mayor cantidad de energía en el espectrograma



6.7 IDENTIFICACION DEL PATRON DE AUDIO

Una vez obtenido el patrón se procede a desplazarlo por toda la señal, muestra a muestra, calculando la correlación cruzada en un instante dado entre el patrón y un segmento de señal de su misma longitud. De esta forma, calculamos en cada instante la similitud entre un trozo de señal y el patrón determinado previamente con poca degradación por el ruido. Lo que quiere decir que el resultado de la comparación presentará valores pequeños en lugares donde el parecido es mayor (zonas donde presumiblemente hay poca degradación) y presentara valores grandes donde el parecido es menor (zonas de una mayor degradación). Este proceso se observa en la figura 40.

Figura 40. Secuencia de identificación de patrón de audio



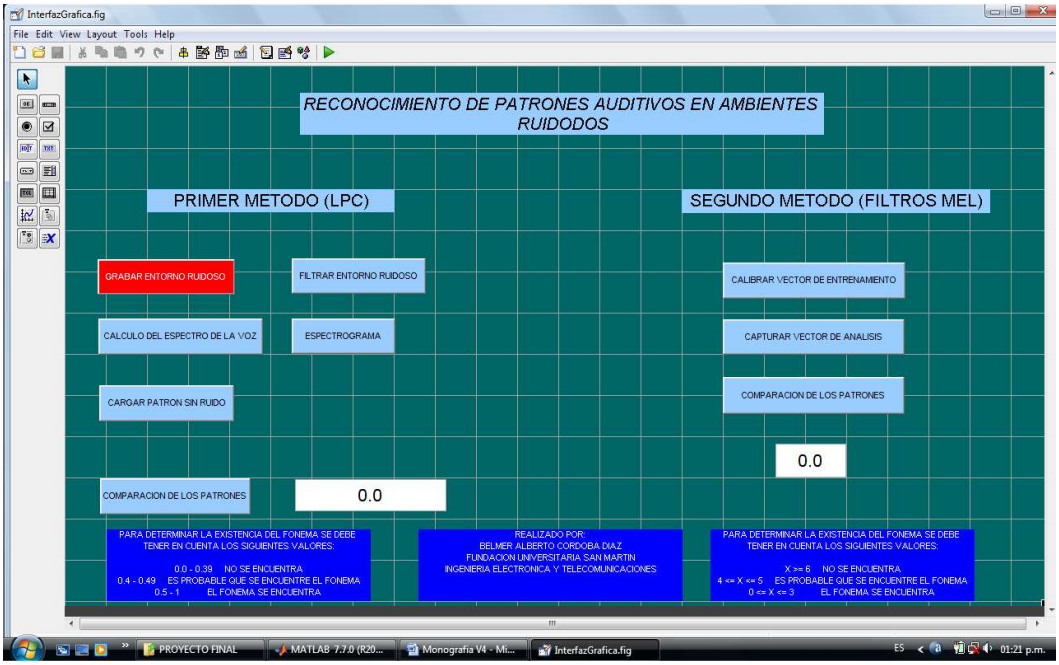
6.8 INTERFAZ GRÁFICA

Para una interacción con el usuario se realizo una interfaz gráfica, la cual se desarrollo por medio de una herramienta de Matlab llamada Guide Quick Start, y que permite realizar una figura interactiva que opera con todos los m-files desarrollados en el proyecto.

Cada uno de los botones permite el desarrollo de una actividad diferente entre las cuales están reflejados a continuación y en la figura 41:

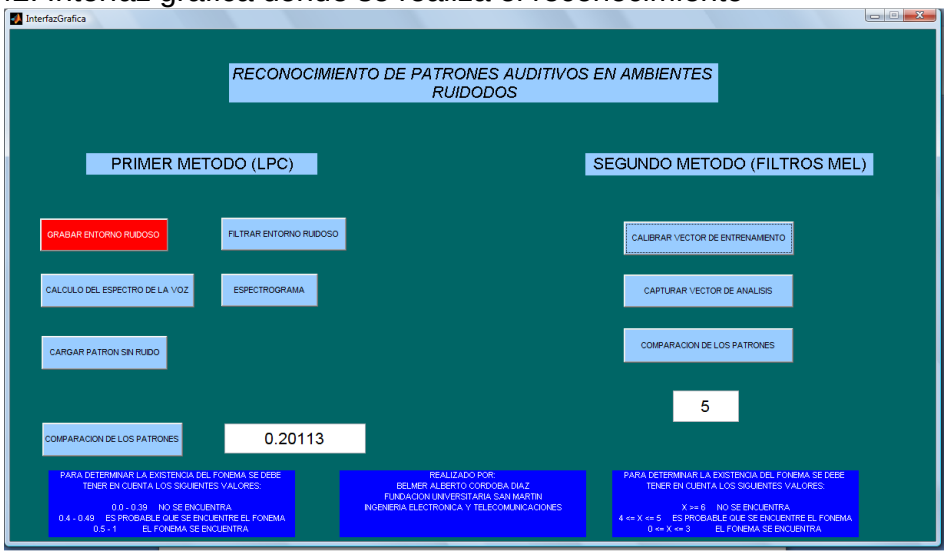
- La grabación de un archivo de audio.
- Graficar el espectro del archivo.
- Generar el espectrograma a partir del cálculo del espectro.
- Ejecutar otra figura o interfaz gráfica en otra ventana distinta.

Figura 41. Ventana de construcción de la interfaz gráfica con herramienta Guide Quick Start



Además en la figura 41 se observa que existen espacio de texto que cambian cada vez que se ejecuta el programa por medio de la interfaz gráfica el cual muestra el resultado de las respectivas comparaciones establecidas en unas variables en los m-files correspondientes para el desarrollo del reconocimiento del patrón de voz. En la figura 42 se observará la interfaz grafica definitiva.

Figura 42. Interfaz gráfica donde se realiza el reconocimiento



7. PRUEBAS Y RESULTADOS

7.1 DIFERENTES CAPTURAS DE VOZ

El procedimiento de pruebas inicialmente cuenta con verificar los diferentes resultados solo cuando se captura la voz de un cierto número de personas. El número de muestras de señales de voz que se muestran a continuación pertenecen a 4 personas distintas (2 hombres y 2 mujeres) en el mismo entorno, es decir con la misma cantidad de ruido degenerativo que pudiesen afectar la señal y pronunciando la misma palabra (Zoológico). Esta palabra fue escogida gracias a que muchos estudios realizados confirman que posee la mayor parte de de los formantes de la voz y gran cantidad de características espectrales. Las mujeres serán representadas como X1 y X2, y los hombres como Y1 y Y2.

La muestra fue tomada para un tiempo de 4 segundos respectivamente. Los resultados fueron los siguientes (figuras 43, 44, 45, 46):

Figura 43. Señal de audio persona X1

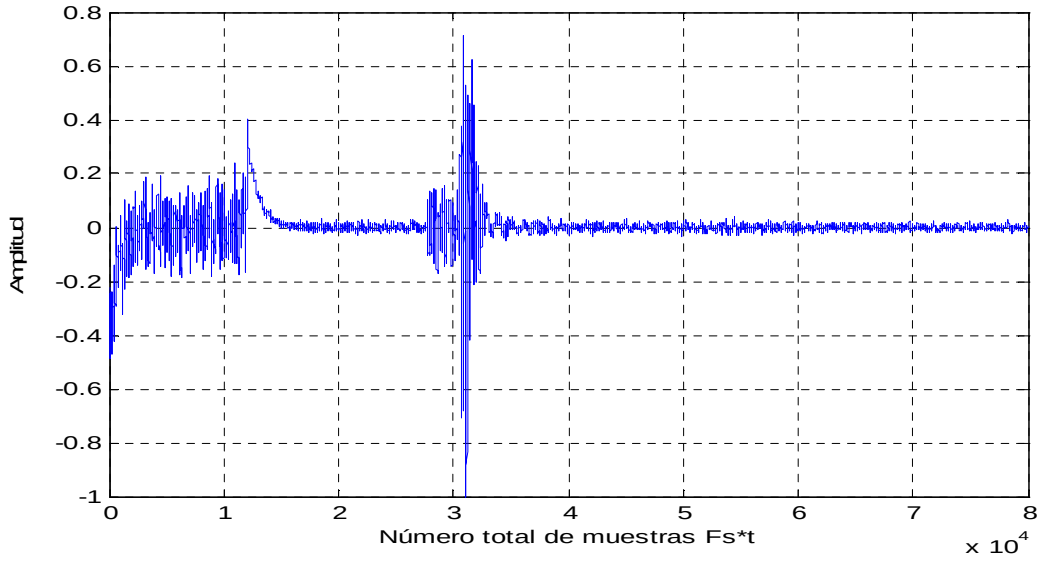


Figura 44. Señal de voz persona X2

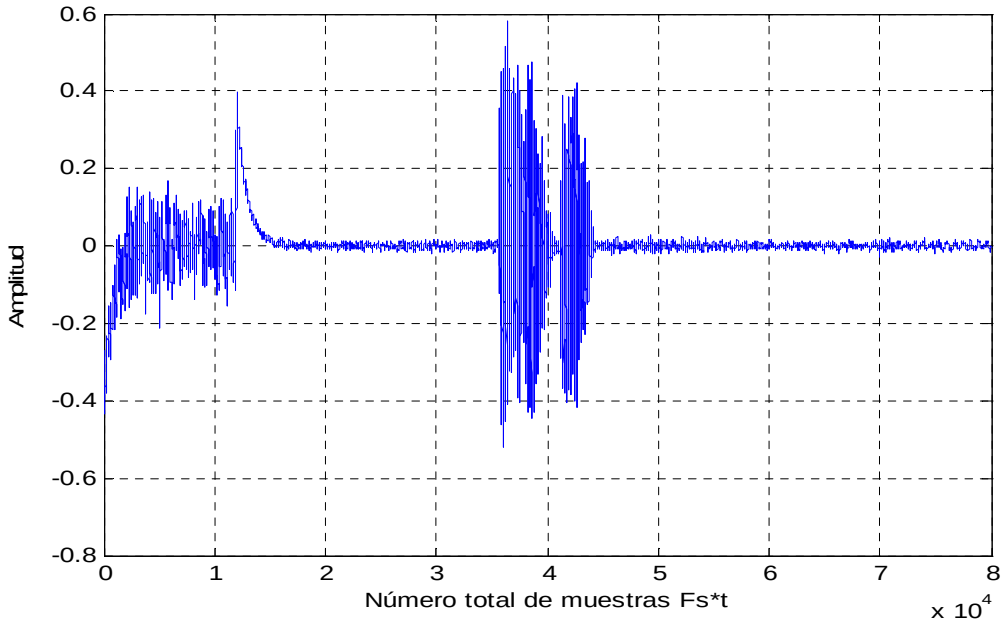


Figura 45. Señal de voz persona Y1

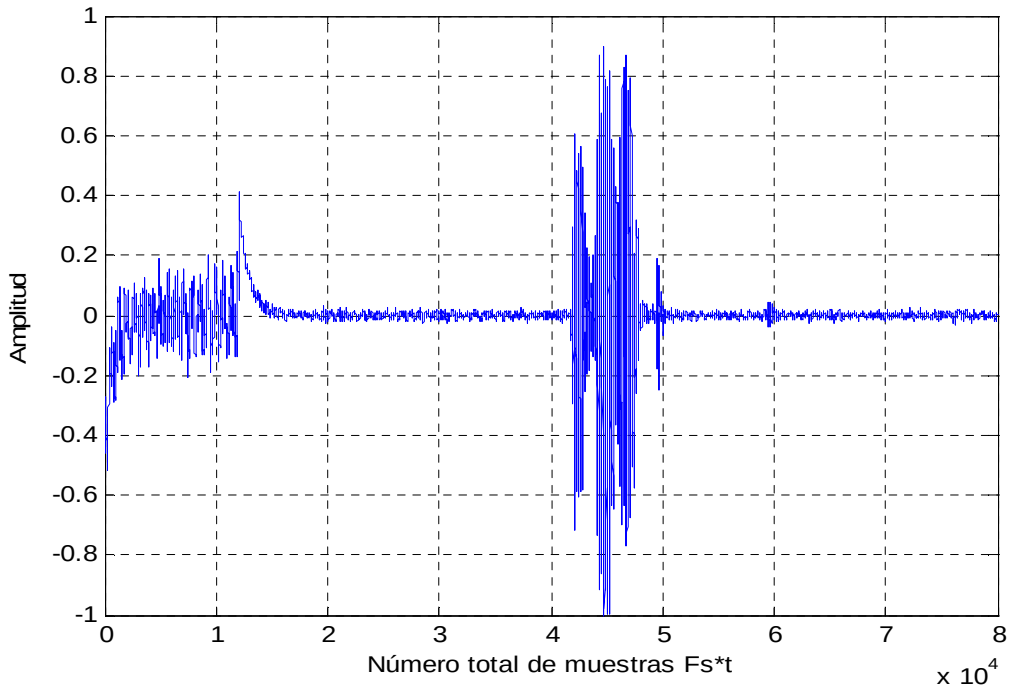
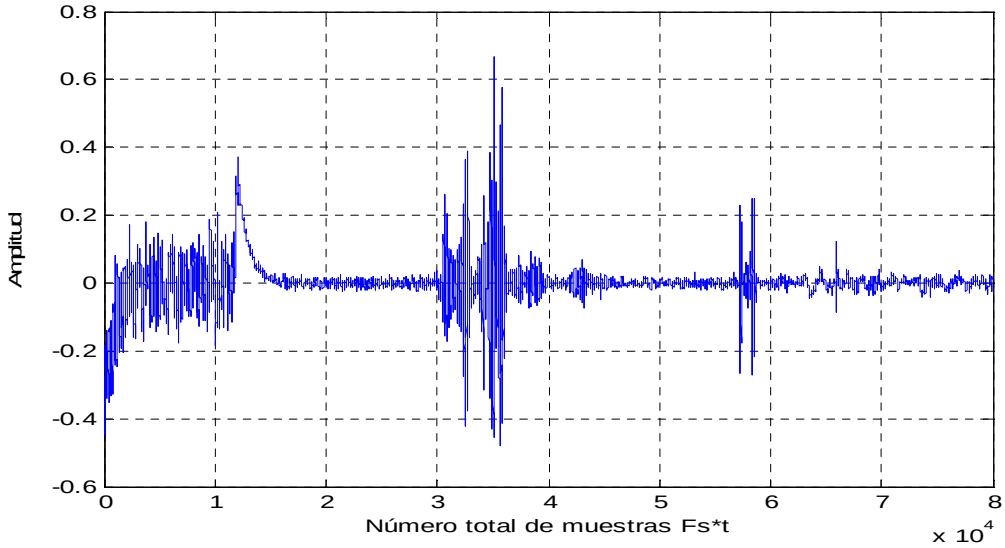


Figura 46. Señal de voz persona Y2

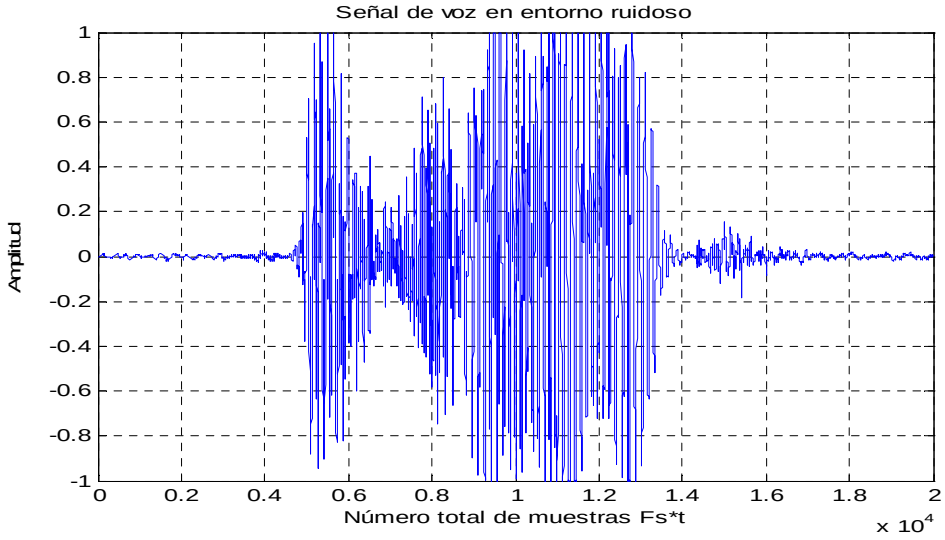


Como se pudo observar todas las cuatro señales poseen espacios de tiempo donde no se presenta gran cantidad de información, además una característica particular la cual se encuentra al inicio del proceso de captura y que puede ser despreciada debido a que es un ruido inherente generado posiblemente por la orden de recopilación de datos.

7.2 ACOTAMIENTO DE LA SEÑAL DE VOZ

Para esta prueba se realiza la reducción del tiempo de muestreo y con la reducción de un número de muestras en la señal la cual ya se encuentra en el dominio del tiempo discreto. Esto con el fin de mantener las muestras con información relevante y además suprimir ese ruido inherente al inicio de la captura. Con un tiempo de captura de 2 segundos y con la eliminación de las 20000 primeras muestras de captura, así la señal resultante queda de representada en un total de 20000 para un total de 1 segundo de muestreo y con la cual se siguió trabajando a lo largo del desarrollo del proyecto es (figura 47):

Figura 47. Señal de voz acotada



7.3 FILTRADO ANÁLOGO Y DIGITAL

Luego de la realización del acotamiento de la señal se procede con el filtrado de esta. En el filtrado análogo existen varios modelos como lo son los propuestos por Butterworth, Chebychev, Bessel, Notch, etc. Además existen distintos tipos de filtrado digital y al igual al filtrado analógico también han sido propuestos distintos métodos conocidos como Eqiripple, Window, Interpoled, etc.

En el desarrollo del proyecto debe decidirse por algún tipo de filtrado y se observo a través de una serie de pruebas de comparación que el filtrado digital es más efectivo que el filtrado analógico. En las figuras 48 y 49 se muestra cada uno de estos representada, como la señal sin filtrar por el color amarillo y la señal filtrada con el color rojizo.

Figura 48. Resultado filtrado análogo

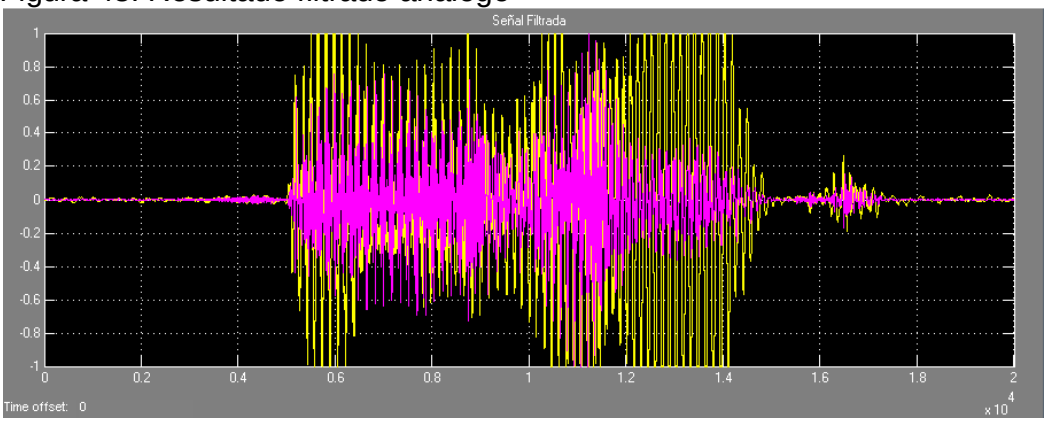
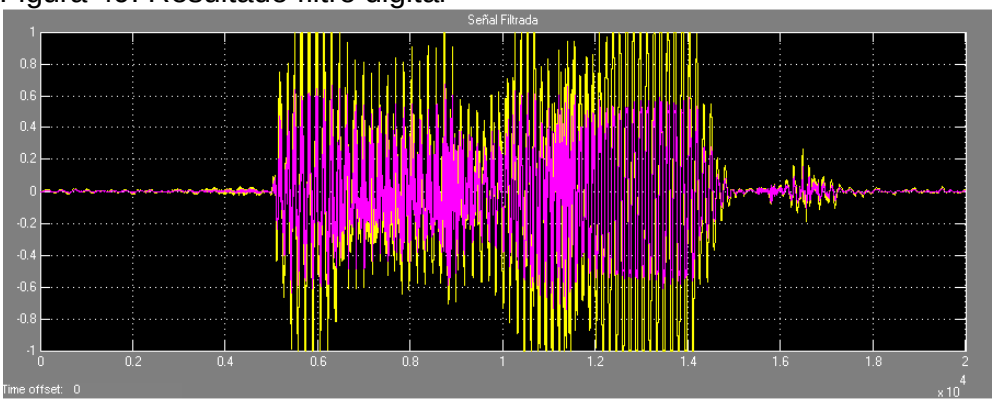


Figura 49. Resultado filtro digital



Aunque el filtrado analógico pareciese a simple vista parece ser mejor que el filtrado digital por lo visto en las figuras 48 y 49, la verdad es que es todo lo contrario esto debido a que en una muestra pequeña no es muy notorio que en algunos sectores de la señal filtrada se presenta un cambio de fase los cuales podrían generar datos erróneos en el espectro de la señal que podrían confundir el resultado del reconocimiento.

Para la identificación de un patrón de audio determinado en la señal es necesario realizar una comparación con otra señal que contenga el patrón por así decirlo, se debe obtener una señal base lo más limpia posible, es decir con el menor número de degradaciones ruidosas posibles. De esta forma la señal base también debe ser filtrada tal y como se muestra en las siguientes figuras 50 y 51.

Figura 50. Filtrado analógico señal base sin ruido

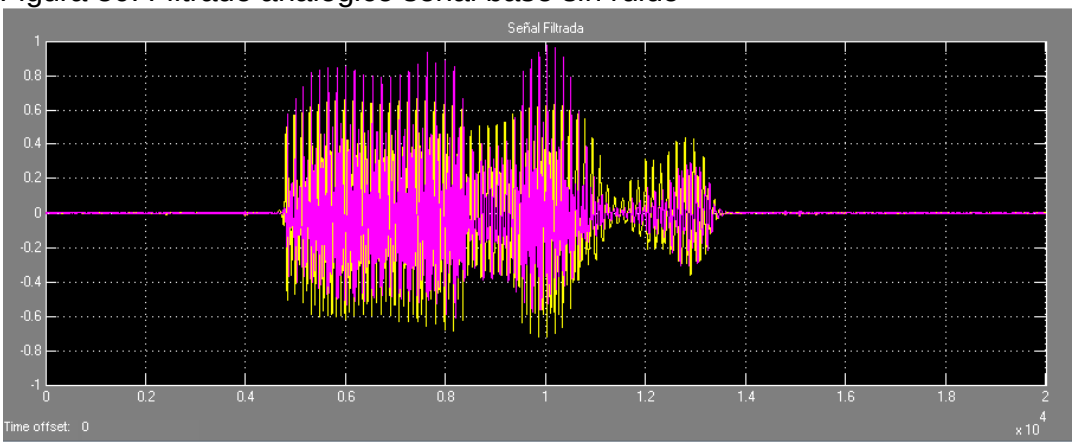
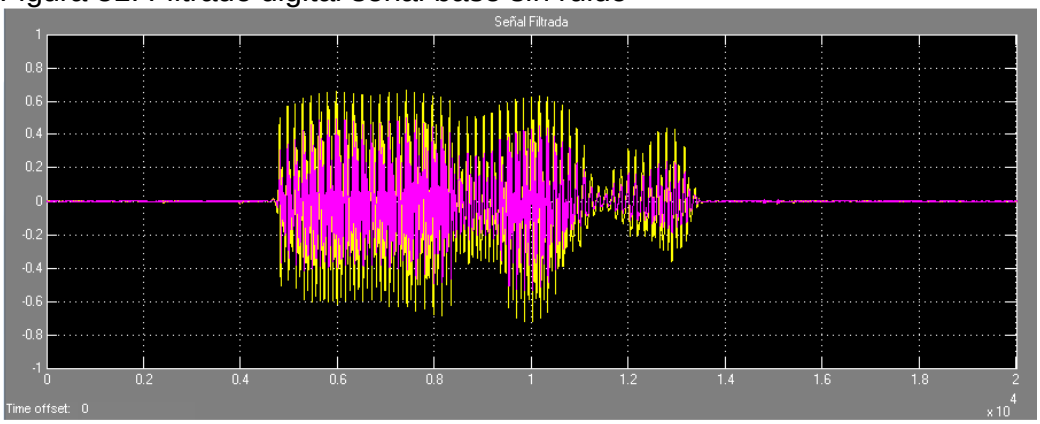


Figura 51. Filtrado digital señal base sin ruido



Como se dijo anteriormente para todas las muestras el filtrado analógico no siempre es el mejor, esto ratifica que el filtrado digital es más efectivo.

7.4 ESPECTRO DE LAS SEÑALES DE AUDIO

El espectro que se obtiene es distinto para cada una de la captura de las señales de audio por ende solo se muestra el espectro de la señal en entorno ruidoso filtrado, esto para dos hombres distintos, los cuales serán representados como se hizo anteriormente por Y1 y Y2.

Figura 52. Señal de voz Y1

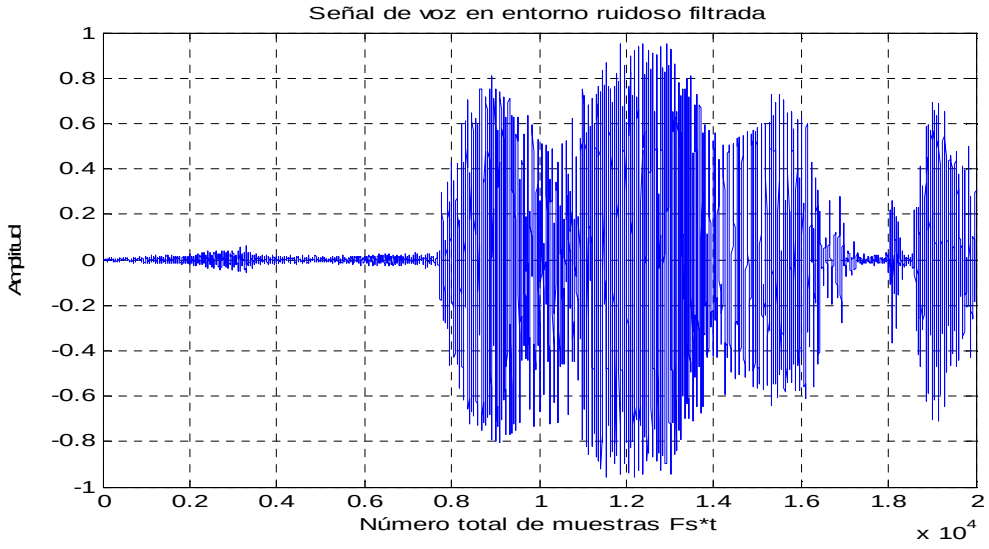


Figura 53. Espectro de la señal de voz persona Y1

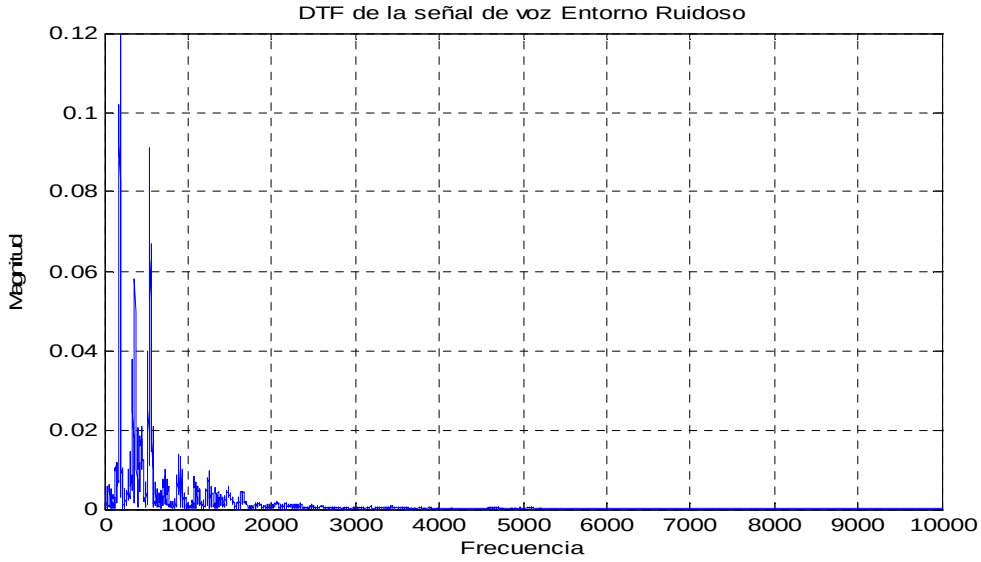


Figura 54. Señal de voz Y2

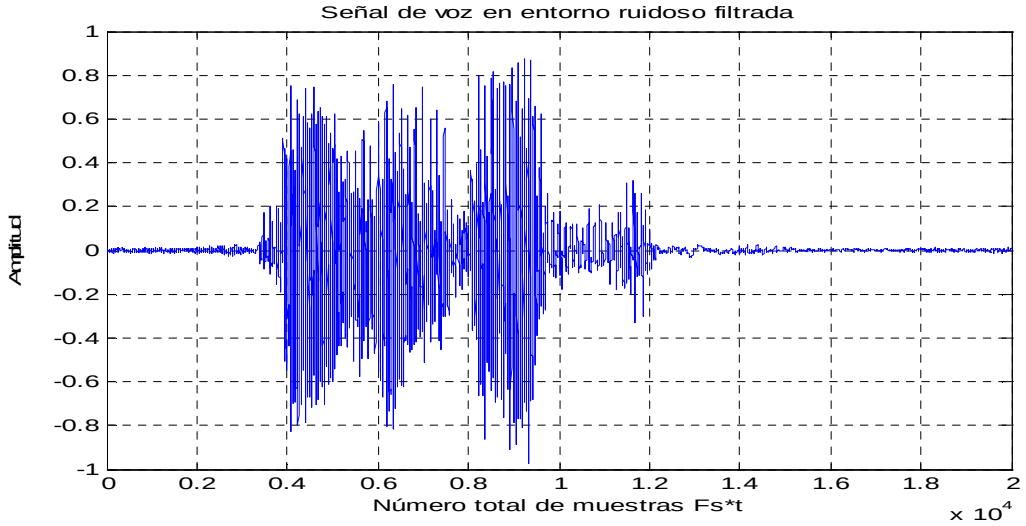
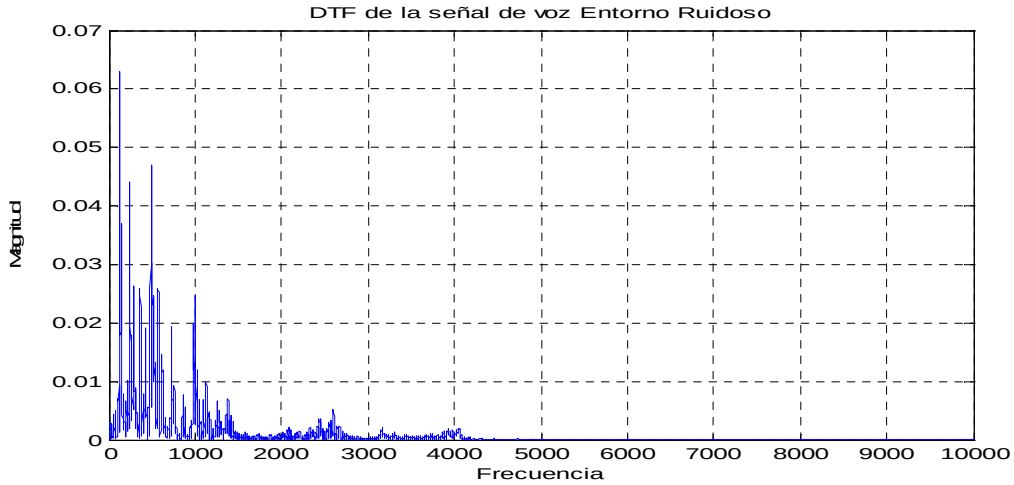


Figura 55. Espectro de la señal de voz persona Y2



Ahora bien, luego de observar los espectros producidos por las dos señales de audio distintas, adicionalmente se graficara el espectro para la señal con la cual se comparara, es decir, la señal capturada previamente y que además fue filtrada (figura 56 y 57).

Figura 56. Señal de voz base filtrada

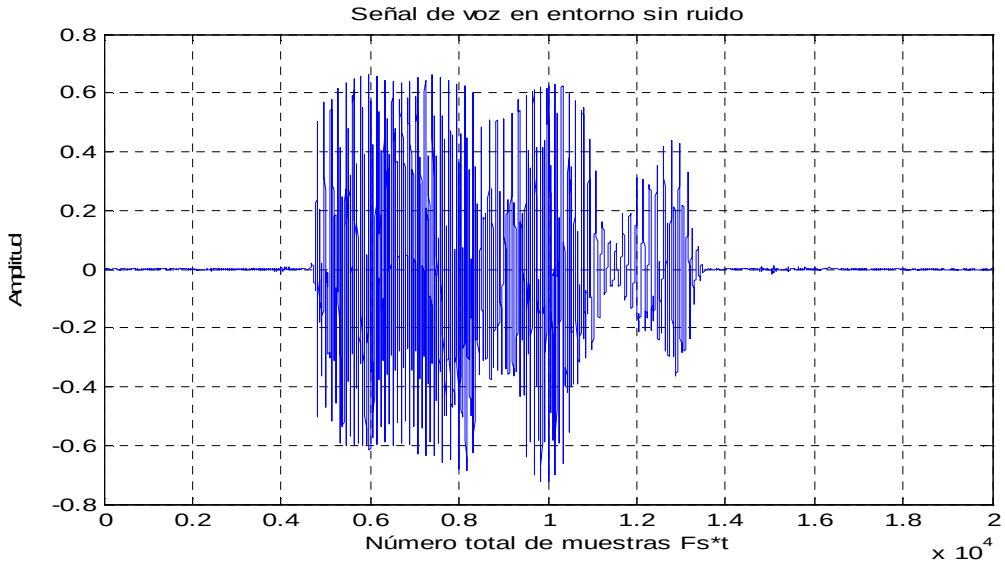
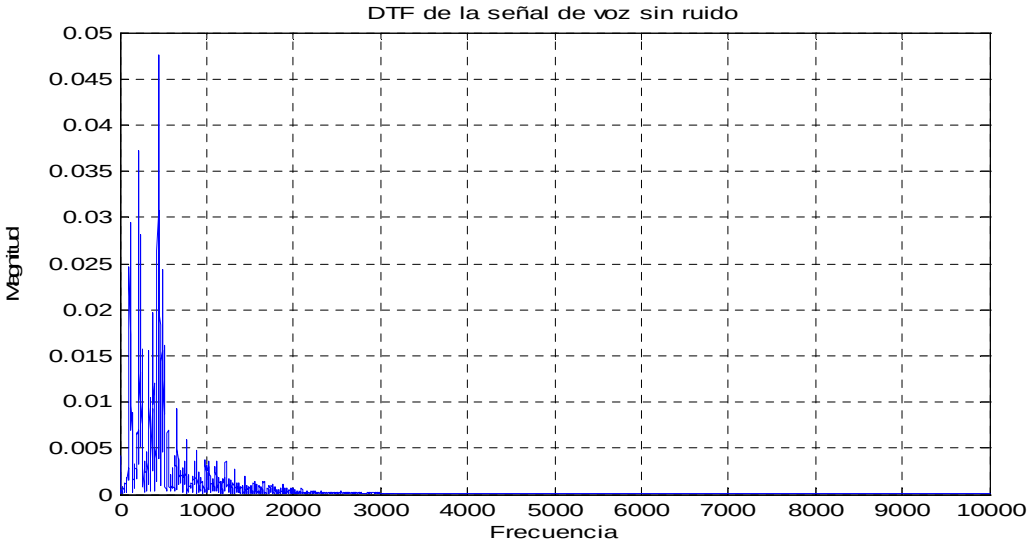


Figura 57. Espectro señal base filtrada



7.5 ESPECTROGRAMAS SEÑALES DE AUDIO

Al igual que en el cálculo del espectro de las respectivas señales se procede a la obtención de los respectivos espectrogramas de las señales de audio (para las personas Y1 y Y2) figuras 58 y 59.

Figura 58. Espectrograma de la señal de voz de persona Y1

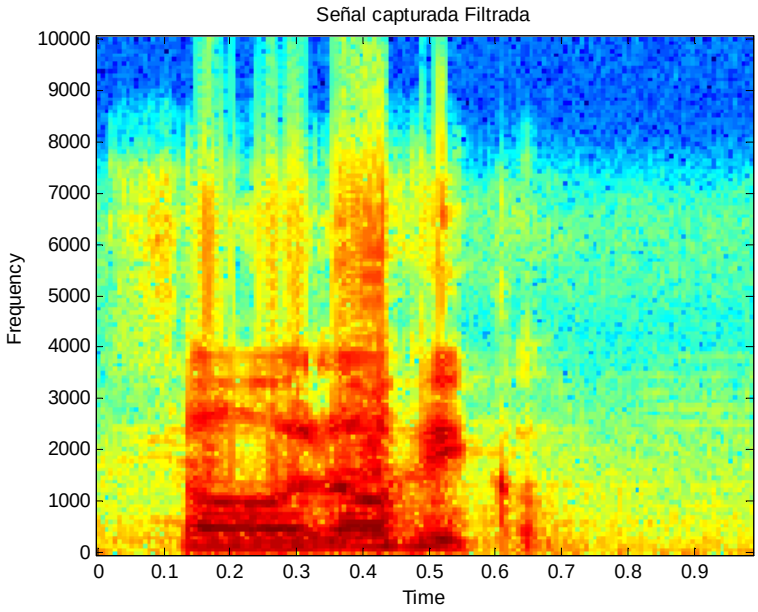
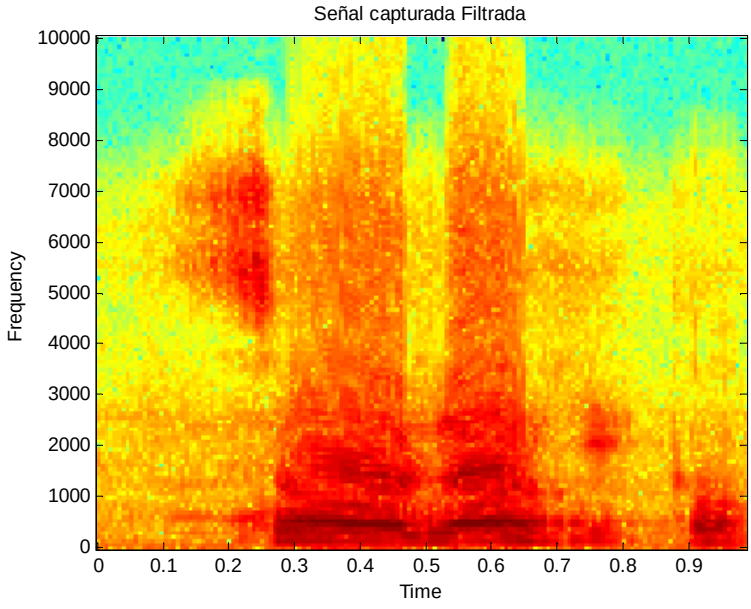
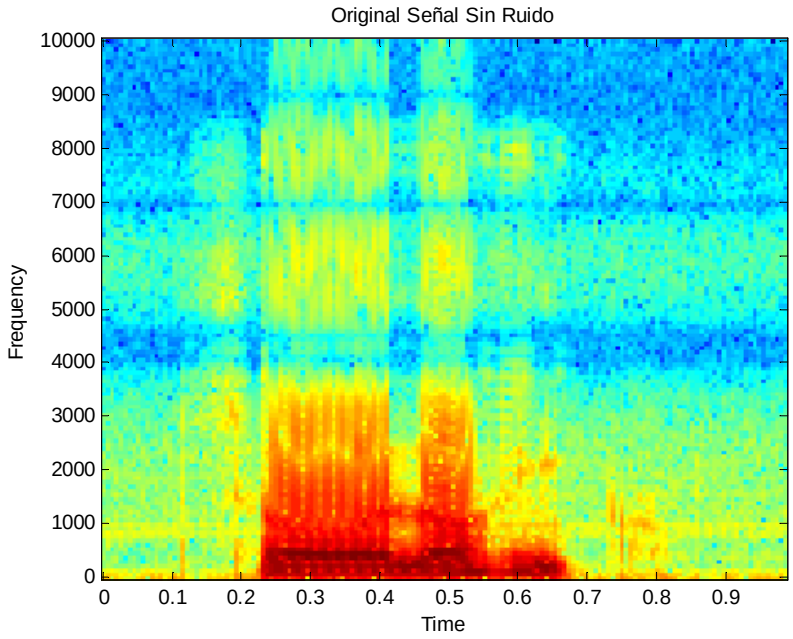


Figura 59. Espectrograma de la señal de voz persona Y2



A continuación en la figura 60 se muestra el espectrograma referente a la señal de voz filtrada en un entorno sin ruido, Con la cual como anteriormente se dijo se comparara, y así luego decir si el patrón se encuentra o no.

Figura 60. Espectrograma señal base filtrada



7.6 PROCESO DE ENTRENAMIENTO

En el proceso de entrenamiento lo que se hace es, tomar varias (10) muestras para así promediarlas, con el fin de obtener un único patrón auditivo característico del fonema, para después compararlo con los patrones fonéticos de las señales de audio en diferentes entornos ruidosos (cafetería universidad, ambientes exteriores y con música de fondo).

Las figuras 61, 62, 63, 64 se muestran cuatro muestras de las 10 en total.

Figura 61. Patrón de entrenamiento 1

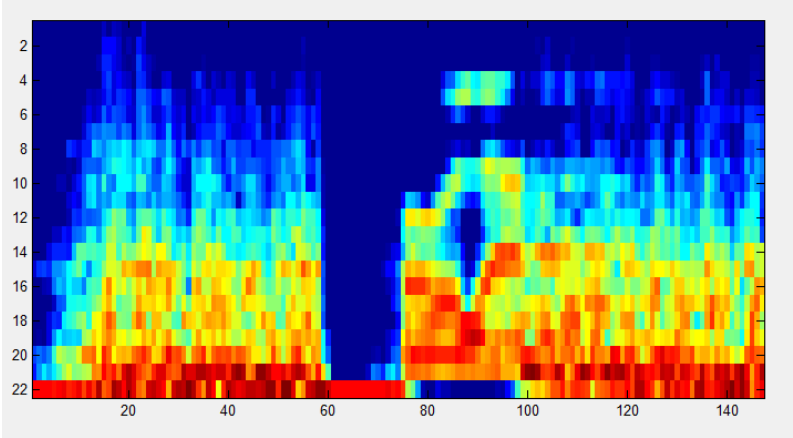


Figura 62. Patrón de entrenamiento 2

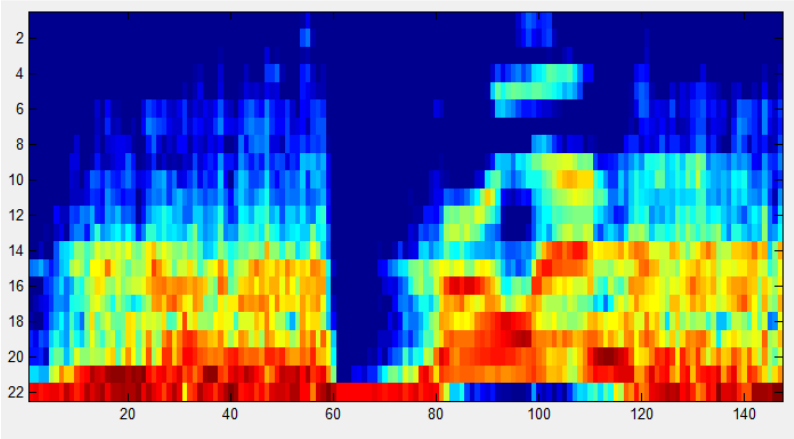


Figura 63. Patrón de entrenamiento 3

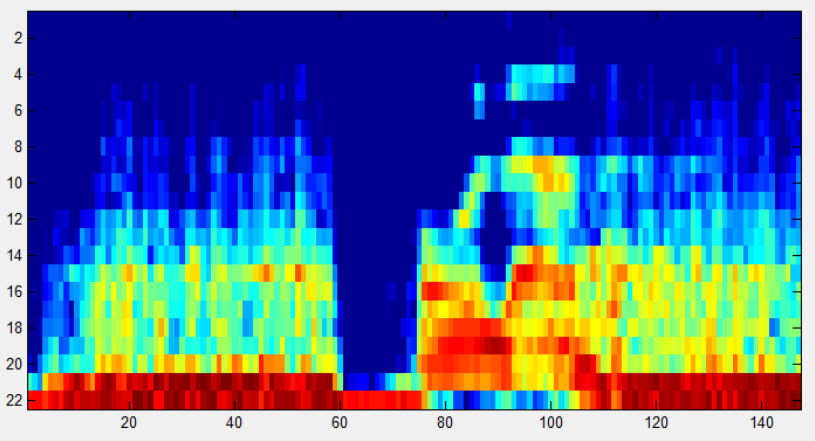
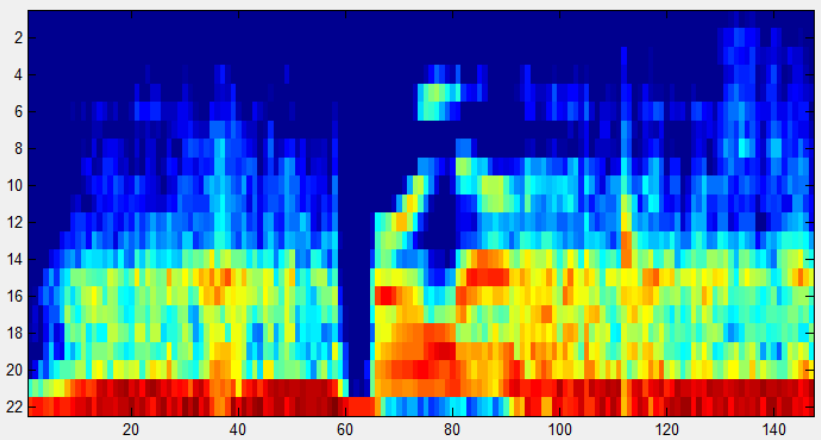


Figura 64. Patrón de entrenamiento 4



Las figuras 65, 66, 67 se representan los espectrogramas de las señales de audio en cada uno de los entornos ruidosos mencionados anteriormente.

Figura 65. Patrón obtenido en cafetería de la universidad

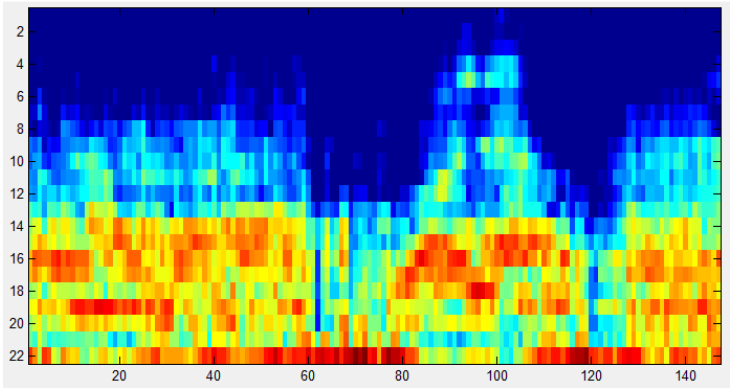


Figura 66. Patrón obtenido en exteriores

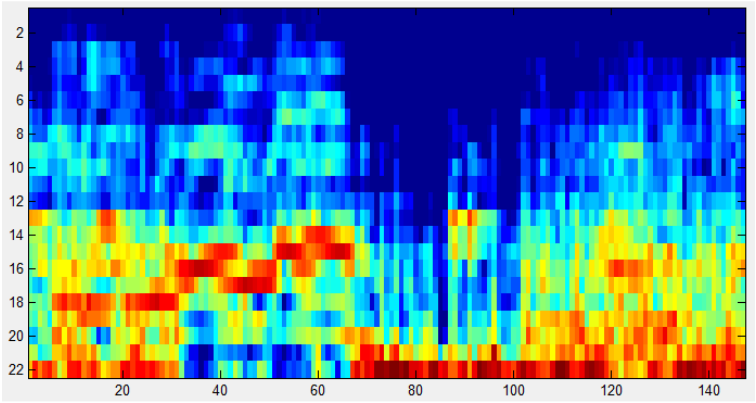
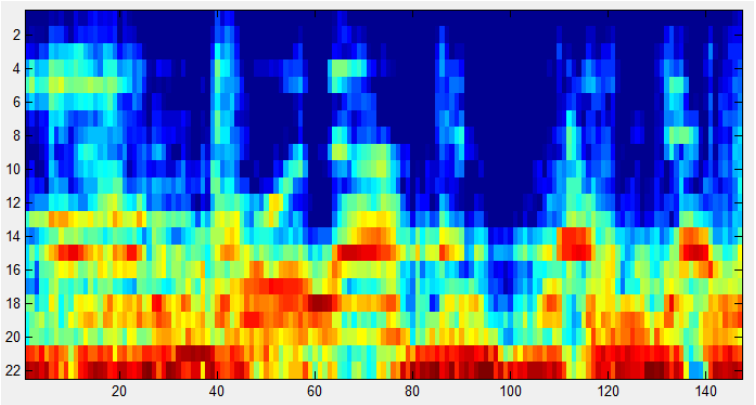


Figura 67. Patrón obtenido con musica de fondo



En la siguiente tabla 2 se relacionan los resultados de las comparaciones hechas de los espectrogramas de los entornos ruidosos con el espectrograma de

entrenamiento. Los resultados que se encuentran entre 0 – 3 garantizan la presencia del patrón fonético, los resultados entre 4 – 5 es posible que se encuentre el patrón fonético y de 6 en adelante no existe la presencia del patrón.

Tabla 2. Resultados de entornos ruidosos

ENTORNO RUIDOSO	PATRON DE ENTRNAMIENTO	RESULTADO
CAFERÍA UNIVERSIDAD		1
EXTERIORES		4
MUSICA DE FONDO		8

8. CONCLUSIONES

- El software de grabación de voz, como el que se encuentra predeterminado por Windows, maneja una extensión para los archivos de audio capturados conocido como .wma (Windows Media Audio). El problema de esta extensión es que ya trae un códec de compresión de audio que genera pérdidas de información, la cual puede ser relevante al momento del cálculo del espectro para la obtención de las formantes características. Por el contrario la extensión .wav (Waveform Audio Format) es la representación con menos pérdidas en la captura de sonidos en el dominio del tiempo discreto.
- La selección de una adecuada frecuencia de muestreo es de vital importancia, ya que de esta depende la cantidad de información que tenga la señal capturada por esto es necesario el cumplir con el teorema de muestreo de Nyquist.
- La implementación de filtros digitales es mucho más efectiva que la implementación de filtros analógicos por dos razones. La primera es que la respuesta transitoria de los filtros análogos es más lenta que la de los filtros digitales y esto se ve reflejado en la respuesta al impulso y la respuesta al paso del filtro digital. La segunda por que los filtros análogos presentan un cambio de fase en algún instante de tiempo perjudicando la información que posea la señal después del filtrado.
- La transformada discreta de Fourier es una herramienta muy necesaria para poder obtener la distribución espectral de energía, debido a que esta permite encontrar en que sectores de la señal de audio en el dominio de la frecuencia se encuentran los coeficientes más representativos o formantes, por medio de los cuales se genera la envolvente espectral.
- Gracias a que la transformada discreta de Fourier a corto plazo se implementa por técnicas de ventaneo, esto permite encontrar los picos representativos de la característica con precisión en cada uno de los segmentos donde se aplicó dicho ventaneo.
- A pesar de que la transformada discreta de Fourier a corto plazo es una buena herramienta para identificar los picos representativos de la característica, esta posee una serie de limitaciones, ya que cuando hay formantes que cambian abruptamente en intervalos cortos de frecuencia la predicción de la envolvente del espectro no es tan efectiva. Por lo que se recomienda implementar otro tipo de herramientas matemáticas que se apliquen al procesamiento digital de señales en el reconocimiento de los patrones de audio.

- Como se trabajo con la transformada discreta de Fourier a corto plazo debido a su fácil implementación, fue necesario la realización de un proceso de entrenamiento, en el cual se promedian las envolventes espectrales de las muestras capturadas para obtener una única envolvente espectral. A diferencia de otro tipo de transformada como wavelet la cual es más compleja de implementar, pero sin la necesidad de la realización de un entrenamiento respectivo.
- Dentro del proceso de decimación del método de reconocimiento de señales de audio conocido como bancos de filtros, se presenta una sobrepocición en los respectivos ventaneos generando perdidas de amplitud y degradaciones de frecuencia.
- Otro método de reconocimiento propuesto pero con deficiencias es el método de Vector de cuantización, en el cual existe una distorsión espectral inherente en el vector de análisis o vector donde se encuentra la muestra de audio en el entorno ruidoso.
- Debido al trabajo con los códigos de predicción lineal LPC los cuales se encargan de la simulación del tracto vocal por medio de un filtro llamado todo polos, fue posible la obtención del espectrograma, el cual es la representación del resultado de calcular el espectro en tramas de una señal, es decir básicamente representar el contenido en frecuencias de la señal conforme varia el tiempo, esta representación esta codificada en colores, donde el color rojo representa la mayor cantidad de densidad espectral de energía.
- Para calcular un espectro de voz exacto es necesario hacer un estudio para determinar el ancho de banda de cada uno de los respectivos ventaneos, para que en este proceso no se pierda algún pico representativo de la característica señal.
- Aunque todos los métodos de reconocimiento de patrones auditivos buscan realizar la comparación de dos vectores característicos (análisis y entrenamiento), la implementación del método de filtros MEL es más efectivo al nivel del reconocimiento, ya que este representa la fusión de dos métodos (bancos de filtros y la variación logarítmica de la escala MEL) haciéndolo más robusto.
- El presente trabajo de grado permitió a su autor una profundización en los diferentes conceptos relacionados con algunos métodos de reconocimiento de patrones de audio, proporcionando el gran reto profesional que conlleva el trabajar en el área del procesamiento digital de señales.

- El ejecutor de este trabajo se siente orgulloso de haber desarrollado un proyecto que apoya el proceso de formación del ingeniero Electrónico y Telecomunicaciones San Martiniano, demostrando la gran importancia que tiene el área del procesamiento digital de señales para el desarrollo del país.

9. RECOMENDACIONES

- Analizar de una forma correcta las formantes del espectro de la señal de audio para así, poder determinar la huella digital que poseen cada uno de los seres humanos en el tracto vocal a través de un espectrograma de frecuencias.
- Debido a que este proyecto se encarga de reconocer palabras, una recomendación es implementar este reconocimiento de fonemas para generar una señal de control en la maniobrabilidad de ciertos dispositivos.
- Gracias a que este proyecto logró reconocer fonemas satisfactoriamente a través de un procesamiento digital de señales donde se implementaron una serie de algoritmos matemáticos, este mismo reconocimiento puede ser realizado a través de un procesador digital de señales (DSP) para lograr su implementación en hardware.
- Para un mejor cálculo de las formantes en la obtención de la envolvente espectral de una señal de audio este puede ser efectuado por medio de análisis Wavelet y autómatas celulares.

GLOSARIO

Amplitud: Es el valor máximo de la Función de Onda y corresponde al máximo valor en voltaje o en decibelios (dB) (Alegsa, 2006) .

Análoga: Forma de expresión de onda o de una señal donde se maneja un rango infinito de puntos en el tiempo (Reference, 2008).

Ancho de banda: Cantidad de información que puede transmitirse en una conexión durante una unidad de tiempo (Gutierres, 2006).

Canal de voz: Canal con un ancho de banda de 300 a 3,400 Hz, indicado para transmisión de voz (Alegsa, 2006).

Fase: Es el valor que en la expresión matemática de la onda toma el argumento de la función (Alegsa, 2006).

Fibra óptica: Tipo de cable que se basa en la transmisión de información por técnicas optoeléctricas. Se caracteriza por un elevado ancho de banda (Gutierres, 2006).

Frecuencia: Número entero de períodos o ciclos alcanzados en la unidad de tiempo por una onda acústica o electromagnética (Reference, 2008).

Interferencia: Superposición de ondas ajenas a la onda que es de interés (Alegsa, 2006).

Perturbación: Es la variación de una magnitud física respecto a un determinado valor que se considera estacionario o de equilibrio (Reference, 2008).

Red: Es un sistema de comunicación de datos que conecta entre sí sistemas informáticos situados en lugares más o menos próximos (Gutierres, 2006).

Señal: Información que se transmite por una red de telecomunicaciones. Puede ser analógica o digital (Alegsa, 2006).

Transductor: es un dispositivo encargado de transformar la naturaleza de la señal (Gutierres, 2006).

Liveness: Tiempo de vida de la señal (Garcia Luz, 2008).

Warmth: Riqueza armónica de la señal por encima de 20KHz (Garcia Luz, 2008).

Reverberación: llegada puntual del retorno de reflexiones tempranas y tardías a punto de generación (Garcia Luz, 2008).

Dereberveración: aislamiento en el tiempo del retorno de reflexiones tempranas y tardías (Garcia Luz, 2008).

Cámara anecoica: cámara libre de reflexiones y refracciones (Garcia Luz, 2008).

BIBLIOGRAFÍA

AB, E. T. (2006). AQM in TEMS Automatic-PESQ Technical Paper. In T. o. b. T. L. M. Ericsson (Eds.)

Alegsa. (Ed.) (2006).

Cadavid, D. (2004). Implementacion de un sistema didactico sobre un canal de voz.

Cadiz, R. (2003). Filtros Digitales. Retrieved Agosto 10, 2009, from http://rodrigocadiz.com/imc/html/Filtros_digitales.html

Dominguez, M. (2001). Señales y Sistemas.

Fernandez Rubio, J. A. (1999). Comunicaciones Analógicas. In E. UPC (Eds.)

Gabiola, F. J. y. A.-H., Basil M. (2007). Análisis y Diseño de Circuitos Analógicos. In E. V. Libros (Eds.)

Garcia Luz, D. L. T. A., BENITEZ Carmen y RUBIO Antonio J. (2008). Speech Recognition, Technologies and Applications. Vienna, Printed in Croatia. In E. B. F. M. a. J. Žibert (Eds.)

Gutierres, F. (2006). Glosario de terminos.

Kioskea. (2007). Introducción de al formato .wav. Retrieved 2 de Agosto, 2009, from http://es.kioskea.net/contents/audio/wav.php3

Lofqvist, R. a. (2006). Haskins Analysis Display and Experiment System. 78

Lopez, E. (2005). Procesamiento Digital de Voz. 68

Marti, M. A. y. L., Joaquim. (1996). Tecnologías del Texto y Habla. In E. U. Barcelona (Eds.)

Media, W. (2008). Windows Media Audio 9. Retrieved 2 Agosto, 2009, from http://www.microsoft.com/windows/windowsmedia/forpros/codecs/audio.aspx

Pianored. (2004). Sonidos wav. Retrieved 2 de Agosto, 2009, from http://www.pianored.com/sonidos-wav.html

QNX. (2008). Retrieved Febrero, 2009, from http://www.qnx.com/news/pr_3156_1.html

Rabiner, L. y. J., Biing-Hwang. (1993). Fundamentals of Speech Recognition. In Prentice-Hall (Eds.)

Reference, W. (Ed.) (2008).

Sancho. (2007). El Espectrograma. Retrieved from http://personal.telefonica.terra.es/web/sixsancal/documentacion/Pdf%20por%20partes/3-2_El%20espectrograma.pdf

Slideshare. (2005). Filtros Digitales. Retrieved agosto 25, 2009, from http://www.slideshare.net/gugaslide/filtros-digitales-presentation

Truong, N. y. S., Gilbert. (1996). Wavelets and Filter Bank. In W.-C. press (Eds.)

Uruguay, F. D. I. M. (2007). Reconocimiento de Voz. Retrieved Septiembre 20, 2009, from http://iie.fing.edu.uy/ense/asign/dsp/proyectos/2001/grupo_b_banco_filtros/Home.htm

Anexo 1

INSTRUCTIVO DE LA INTERFAZ GRAFICA (RECONOCIMIENTO DE PATRONES AUDITIVOS EN AMBIENTES RUIDOSOS)

1. Este proyecto fue desarrollado en un software matemático llamado Matlab r2008b.
2. Ubicar el m-file llamado InterfazGrafica en la misma ubicación con todos los demás archivos que de los cuales la interfaz grafica dependa.
3. Abrir el archivo InterfazGarfica.m y ejecutarlo dando click en run, luego se despliega la ventana InterfazGarfica.fig.
4. En la interfaz grafica encontrara dos métodos de reconocimiento LPC y filtros MEL.
5. Como operar con LPC:
   - 5.1 Grabar un archivo de audio en entorno ruidoso (2 segundos de tiempo de captura), se graficara la señal de voz en entorno ruidoso capturada.
   - 5.2 Filtrar la señal en un entorno ruidoso, se graficara la señal de voz en entorno ruidoso filtrada.
   - 5.3 Cálculo del espectro de la voz, se graficara la DTF de la señal de voz en entorno ruidoso.
   - 5.4 Generar el espectrograma, se graficara el espectrograma de la señal del entorno ruidoso filtrada.
   - 5.5 Cargar patrón sin ruido, se mostraran tres graficas ya almacenadas previamente en memoria (señal de voz en entorno sin ruido filtrada, la DTF de la señal de voz sin ruido y el espectrograma de la señal del entorno sin ruido).
   - 5.6 Comparación de los patrones, para ver el resultado de la comparación dar Enter en el espacio en blanco y así cargarlo.
   - 5.7 Los intervalos para interpretar el reconocimiento fonético por el método LPC están dados en la parte inferior izquierda.
6. Como operar con filtros MEL:
   - 6.1 Calibrar vector de entrenamiento, aparecerá otra ventana llamada recordtool.fig.
     - 6.1.1 Grabar, captura la muestra de audio y genera el espectrograma.
     - 6.1.2 Salvar, salva la muestra capturada.
     - 6.1.3 Cargar, carga el espectrograma de la muestra ya guardada.
     - 6.1.4 Lista de muestras, en la parte inferior izquierda indica con que muestra se está trabajando, hay que tomar diez muestras.
   - 6.2 Capturar vector de análisis, grabara la señal de voz en entorno ruidoso.
   - 6.3 Comparación de los patrones, para ver el resultado de la comparación dar Enter en el espacio en blanco y así cargarlo.
   - 6.4 Los intervalos para interpretar el reconocimiento fonético por el método filtros MEL están dados en la parte inferior derecha.