ESTIMACIÓN DE LA FRECUENCIA FUNDAMENTAL DE SEÑALES DE VOZ DEL SUROCCIDENTE COLOMBIANO APLICANDO LA TÉCNICA *WAVELET*

Bairon Herney Alvira Manios
Andrés Eduardo Sarria Manrique

Director: Ing. Harold A. Romo R.

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telecomunicaciones
Popayán, noviembre de 2007

AGRADECIMIENTOS

*Gracias a **Dios** por protegerme y brindarme la oportunidad de educarme y formarme como profesional y como ser humano.*

Todo lo que soy y lo que he logrado hasta hoy no hubiera sido posible sin el apoyo de mi familia. Este trabajo esta dedicado:

A mi madre Maria Luisa Manios y a mis hermanos: Javier, Andrés, Darío y Leidy Alvira Manios por ser el soporte de mis metas, por su apoyo, fortaleza y carácter a la hora de formarme; a mi abuela Sinforosa Quimbayo y a mis tíos Gilberto, Arturo, Cecilia, Vicente, Socorro, Elías y Félix Manios Quimbayo por creer en mi; a mi primita Laura Sofía; al resto de la familia y a mi novia Lili Lucia Meneses y sus padres por su apoyo.

A quienes hicieron parte de mi formación profesional profesores e ingenieros. Agradezco al Ingeniero Harold A. Romo R., director del proyecto por su paciencia y ejemplo quien no solamente me formo en lo profesional sino también en lo personal y humano; a la Doctora Miryam Adela Barreto por su ayuda en la realización de este proyecto; a mi compañero de tesis Andrés Eduardo Sarria por su amistad y paciencia, a Clara Fernández por su valiosa colaboración y a todos los demás profesores y compañeros de estudio que hicieron parte de mi formación.

Finalmente agradezco A mis leales amigos Nicolás, Jairo, Luís Alberto, Cristina, Iván, Pastor, Efrén con quienes compartí momentos inolvidables, a Rosa, y a todas aquellas personas que hicieron y son parte de mi vida.

Bairon.

*Gracias quiero darle al buen **Dios** por tomarme de la mano y ser la luz que orienta mi vida, gracias por el regalo de la **Santísima y siempre Virgen María** quien con su maternal amor aboga siempre por mí ante nuestro salvador. A Ellos este trabajo, y toda mi vida.*

Este pequeño hijo de Dios, da gracias a quienes con su amor, paciencia, espera y confianza, me han ayudado a formarme y que de seguro este trabajo, sin ellos no sería posible:

A mis papás: Hugo Sarria y Leidy Manrique, a mis hermanos: Anderson y Paula, a ellos dedico este trabajo y todo mi esfuerzo durante la carrera; A mis abuelos, tíos, primos, a mi lindo sobrino y todos mis familiares, gracias por creer en mi; a mi Clarita, un gracias grandote, gracias por tu amor, ayuda y trasnocho, ha sido muy valiosa tu compañía y colaboración; A todas la familias que en este hermosa ciudad Blanca me acogieron, gracias por el calor de hogar.

Agradezco a todos aquellos sin duda importantes por su incondicional apoyo,

Al director de tesis: Ing. Harold Romo, gracias por su espera, alegría y ejemplo, sus orientaciones nos ayudaron en la tesis y servirán para nuestra vida; a la Fonoaudióloga Miryam Adela Barreto, por su colaboración y por ayudarnos a entender muchos de los temas desarrollados; a mi compañero de tesis Bairon Herney Alvira, por las chocolatadas, las ponys, el pan con mortadela y por su buena y chévere amistad ofrecida en estos años de anteproyectos y de la tesis; a todos mis profesores de la "U", compañeros y amigos de clase, a los de la Casa#19, sé que Dios los puso en mi camino para mostrarme lo bondadoso que eran cada uno de ustedes; A la maravillosa Universidad del Cauca, a la Red de Datos, por todo lo que en mi han grabado; a los grupos de oración, a mi Santa Iglesia Católica, gracias por ayudarme a ser cada día mejor.

A todos mis más sinceros agradecimientos.

Andrés.

GLOSARIO DE TÉRMINOS

ANAMNESIS Conjunto de los datos clínicos relevantes y otros del historial de un paciente. || Reminiscencia (‖ representación o traída a la memoria de algo pasado).

CGAWAVF Complex Gaussian Adapted *Wavelet* Family – Familia *Wavelet* Compleja Gaussiana Adaptada.

DFT Transformada Discreta de Fourier.

FF Frecuencia Fundamental.

FFT Fast Fourier Transform – Transformada rápida de Fourier.

IDTF Transformada Discreta de Fourier Inversa.

LPC Codificación por Predicción lineal.

LSP Lines Spectral Pair – Pares de Lineas Espectrales.

SFS Speech Filing System.

TWC Transformada *Wavelet* Continua.

TWD Transformada *Wavelet* Discreta.

CONTENIDO

INTRODUCCIÓN....8

1. PRODUCCIÓN DE LA VOZ Y LAS TÉCNICAS CLÁSICAS DE TRATAMIENTO DE SEÑALES DE VOZ....10

1.1 LA VOZ HUMANA....10

1.2 EL APARATO VOCAL: ANATOMÍA Y FISIOLOGÍA....11
1.2.1 Cavidades infraglóticas (Órganos respiratórios) [1][2]....12
1.2.2 La cavidad glótica [1][2]....12
1.2.3 Las cavidades supraglóticas [1][2]....14

1.3 PRODUCCIÓN ACÚSTICA DEL HABLA....16
1.3.1 Sonidos vocálicos....19

1.4 TÉCNICAS DE ANÁLISIS....20
1.4.1 Transformada Discreta de Fourier (DFT)....20
1.4.2 Técnica de análisis con cruces por cero....21
1.4.3 Autocorrelación....22
1.4.4 Análisis Cepstral....23
1.4.5 Banco de filtros....24
1.4.6 Codificación por predicción lineal (LPC)....25

2. TÉCNICA *WAVELET* EN EL PROCESAMIENTO DE SEÑALES DE VOZ....27

2.1. *WAVELETS*....27

2.2 TRANSFORMADA *WAVELET*....31
2.2.1 Transformada *wavelet* continua....31
2.2.2 Transformada *wavelet* Discreta....32

2.3 FAMILIA *WAVELET* PARA EL TRATAMIENTO DE LAS SEÑALES DE VOZ....33
2.3.1 Selección de la *wavelet* madre....34
2.3.2 *Wavelet* Gaussiana Compleja....36

2.4 EL OÍDO HUMANO COMO MODELO DE ANÁLISIS....37
2.4.1 Escala de Bark....39
2.4.2 Coeficientes adaptados....41

2.5 CREACIÓN DE LA FAMILIA *WAVELET* CGAWAVF....41

2.6 ESTRUCTURAS PARA EL PROCESAMIENTO DE LAS SEÑALES DE VOZ....43
2.6.1 *Wavelets* diádicas....43
2.6.2 Bancos de filtros....44
2.6.3 Estructura de bandas adaptadas.....44

3. ALGORITMO DE ESTIMACIÓN DE LA FRECUENCIA FUNDAMENTAL DE SEÑALES DE VOZ CON LA FAMILIA WAVELET CGAWAVF....47

3.1 FASE DE ANÁLISIS 47

3.2 FASE DE DISEÑO 48
3.2.1 Toolbox *wavelet* y entorno de desarrollo de interfaces gráficas de usuario 48
3.2.2 Diagrama en bloques del sistema 48
3.2.2.1 Adquisición 49
3.2.2.2 Pre-procesamiento 51
3.2.2.3 Procesamiento wavelet 53
3.2.2.4 Análisis 59
3.2.2.5 Estimación 64

3.3 FASE DE IMPLEMENTACIÓN 65
3.3.1 Funciones de procesamiento 65
3.3.2 Funciones de análisis 67

3.4 CARACTERÍSTICAS 69

3.5 MODO DE EJECUCIÓN DEL ALGORITMO 69

3.6 LIMITACIONES 70

4. SISTEMA CGAWAVF, LA VOCUDC Y EL SPEECH FILING SYSTEM 71

4.1 BASE DE DATOS DE VOCES DEL SUROCCIDENTE COLOMBIANO 71
4.1.1 Anamnesis de Voz [1][2] 72
4.1.1.1 Plantilla de la Encuesta para la valoración de una voz normal: 73
4.1.1.2 Los hablantes: Población muestra. 78
4.1.1.3 La grabación de la voz 79
4.1.1.4 Implementación de la base de datos 81
4.1.1.5 Administración de la base de datos 82
4.1.1.6 Formas de Ondas 83

4.2 SPEECH FILING SYSTEM (SFS). 84
4.2.1 Reseña del Speech Filing System 85
4.2.2 Estimación de la *FF* con SFS. 85

4.3 RESULTADOS. 87

5. CONCLUSIONES Y RECOMENDACIONES 91

BIBLIOGRAFÍA 95

ÍNDICE DE FIGURAS

Figura 1.1 Esquema en conjunto del aparato vocal 11
Figura 1.2 Esquema de la laringe 13
Figura 1.4 Corte transversal de la laringe. Movimiento del cartílago aritenoides y de los pliegues vocales (líneas continuas o discontinuas) 13
Figura 1.5 Faringe, la cavidad bucal y la cavidad nasal 15
Figura 1.6 Onda glotal compleja en el dominio de la frecuencia 17
Figura 1.7 Onda glotal compleja en el dominio del tiempo 17
Figura 1.8 Señal generada por la fuente plosiva. 18
Figura 1.9 Análisis Cepstral 23
Figura 2.1 Zoom en la *wavelet* Daubechies3. 29
Figura 2.2 Traslación y dilatación de una función *wavelet* madre. 32
Figura 2.3 Recubrimiento en el plano *wavelet.* 33
Figura 2.4 Función Gaussiana. 37
Figura 2.5 *Wavelet* compleja gaussiana de orden 8. (a) Parte real. (b) Parte Imaginaria. 37
Figura 2.6 Partes del sistema auditivo humano. 38
Figura 2.7 Estiramiento de la Cóclea: (a) Corte longitudinal y (b) Corte transversal. 38
Figura 2.8 Respuesta a frecuencias de la membrana Basilar. 39
Figura 2.9 Escala de Bark y la membrana Basilar. Respuesta a frecuencias: altas (a) y bajas (b). 39
Figura 2.10 *Wavelet* de la Banda 1, parámetros a = 2700X10^{-6} y w_0 =0.8482. 42
Figura 2.11 Espectro frecuencial (parte real) de las 17 *wavelet*s de la nueva familia. 43
Figura 2.12 Modulo de la Transformada de Fourier de las 17 43
Figura 2.13 (a) Descomposición *wavelet* diádica. (b) Bancos de Filtros de dos canales 44
Figura 2.14 Esquema de análisis mediante la familia *wavelet cgawavf.* 45
Figura 3.1 Diagrama de bloques del Sistema CGAWAVF 49
Figura 3.2 Diagrama de flujo del módulo de *Adquisición*. 49
Figura 3.3 Grafica de un registro de voz (vocal a) en la base de datos VOCUDC 50
Figura 3.4 Espectro de frecuencia de la señal (vocal a). 51
Figura 3.5 Diagrama de flujo del módulo de *Pre-procesamiento* 51
Figura 3.6 Representación en el dominio del tiempo del recorte de las colas de la señal 52
Figura 3.7 Diagrama de flujo del módulo de *Procesamiento Wavelet*. 53
Figura 3.8 *Wavelet* de la Banda 9 (a) Parte Real (b) Parte Imaginaria 55
Figura 3.9 Procesamiento de la señal de entrada con la parte real de la *wavelet* de la banda 9. 56
Figura 3.10 Procesamiento de la señal de entrada con la parte imaginaria de la wavelet de la banda 9. 57
Figura 3.11 Gráfica de la magnitud de la señal procesada por la *wavelet* de la banda 9 58
Figura 3.12 Diagrama de flujo del módulo de *Análisis*. 59

Figura 3.13 Grafica de máximos y PRs en un segmento de la señal procesada por la wavelet de la banda 9. 61
Figura 3.14 Grafica de análisis cruzado de bandas tomando como referencia a FR_j1, del vector $V_{FR}1$ 63
Figura 3.15 Diagrama de flujo del módulo de *Estimación*. 64
Figura 4.1 Micrófono Sony c-48. 80
Figura 4.2 Tarjeta de sonido Creative E-MU 1820. 80
Figura 4.3 Selección del inicio y final de la vocal 82
Figura 4.4 Pistas de las vocales divididas 82
Figura 4.5 Vocal /a/ femenina: (a) intervalo de 400ms. (b) intervalo de 20 ms. 83
Figura 4.6 Vocal /i/ femenina (a) intervalo de 400ms. (b) intervalo de 20 ms. 83
Figura 4.7 Vocal /a/ masculina (a) intervalo de 400ms. (b) intervalo de 20 ms. 84
Figura 4.8 Vocal /i/ masculina (a) intervalo de 400ms. (b) intervalo de 20 ms. 84
Figura 4.9 Archivo de Audio en SFS. 85
Figura 4.10 Estimación de la frecuencia fundamental con SFS. 86
Figura 4.11 Ítems de la señal de voz y la estimación de la frecuencia fundamental 86
Figura 4.12 Matriz de valores estimados de la frecuencia fundamental 87
Figura 4.13 Valores Promedios en hombres SFS Vs CGAWAVF 89
Figura 4.14 Valores Promedios en hombres SFS Vs CGAWAVF 89

ÍNDICE DE TABLAS

Tabla 2.1 Indicaciones de la regularidad de las *wavelet*s Daubechies.....30
Tabla 2.2 Valores para cada banda critica en la en la Escala de Bark 40
Tabla 2.3 Valores de los coeficientes adaptados.....41
Tabla 3.1 Características de las *wavelets* asociadas a cada una de las bandas 54
Tabla 3.2 Número de muestras para cada una de las wavelets.....55
Tabla 4.1 Edades de los hablantes.....79
Tabla 4.2 Valores promedios del SFS y CGAWAVF.....88
Tabla 4.3 Valores de la Desviación estándar del SFS y CGAWAVF.....88
Tabla 4.4 Características acústicas de las vocales del español rioplatense 90

INTRODUCCIÓN

Hoy en día el avance de la ciencia exige como prioridad que la investigación se oriente de manera interdisciplinaria, cualidad que permite el desarrollo y aplicación de fundamentos, análisis y modelos que respondan lo más acertadamente posible con la realidad; de este modo, tenemos en el gran universo de la investigación aplicada la preocupación por conocer el proceso de la voz[1], que tiene una gran aplicabilidad no sólo en los diferentes campos de la medicina sino también en el campo de la ingeniería como es nuestro caso.

En este sentido, el procesamiento de señales por medio de la técnica *wavelet* es el área de la ingeniería en que se desarrolló este trabajo; en la Universidad del Cauca se han aplicado estas técnicas para el tratamiento de imágenes y señales electromiográficas entre otras, pero la aplicación de ésta en señales de voz es poco estudiada, lo cual fue uno de los motivos para tratarlas, además de la importancia de poder generar un aporte a la reciente preocupación por los problemas que implican la voz.

El proyecto “Estimación de la frecuencia fundamental de señales de voz del suroccidente colombiano aplicando la técnica wavelet”, tiene su origen en el tratamiento digital de señales y particularmente en la caracterización de la frecuencia fundamental FF[2] de señales de voz, lo cual implica tener como punto de partida un conocimiento básico sobre el proceso fisiológico normal de la voz, lo que facilita la comprensión en los diversos procesos desarrollados. Brevemente se presentarán algunas de las técnicas clásicas en el tratamiento de señales de voz para la estimación de la frecuencia fundamental, algunas de ellas basadas en la teoría de Fourier; se expondrá la teoría *wavelet* y sus aplicaciones en el procesamiento de señales de voz, identificando las familias wavelet más adecuadas para este tipo de señales; además se presenta el modo

[1] La voz son ondas sonoras producidas en la laringe por la salida del aire (espiración) que, al atravesar las cuerdas vocales, las hace vibrar, en "Lenguaje y alteraciones del lenguaje." *Microsoft Encarta* 2006.

[2] La frecuencia fundamental FF es el nivel óptimo en el cual la voz produce una frecuencia confortable sin la menor tensión laríngea y sin esfuerzo [1] p.169.

de aplicación del algoritmo de Leonard J. García [8] al cual se le realizaron algunas modificaciones pertinentes para el procesamiento de las señales de voz almacenadas en la base de datos VocUDC desarrollada para este proyecto[3] simulando y evaluando el desempeño del algoritmo con respecto a referencias bibliográficas y a la herramienta software "*Speech Filing System*".

[3] Se implementaron, crearon y modificaron otras funciones que junto con la interfaz gráfica permiten verificar y contrastar los resultados en la estimación de la frecuencia fundamental como las funciones del toolbox de matlab por ejemplo.

1. PRODUCCIÓN DE LA VOZ Y LAS TÉCNICAS CLÁSICAS DE TRATAMIENTO DE SEÑALES DE VOZ

La ciencia que estudia el sonido es conocida como acústica, fenómeno mecánico que tiene sus fundamentos en la física y especialmente en las leyes del movimiento, en ella se abre paso el estudio de los sonidos de la voz humana, área que conocemos como *fonética acústica.* Por ende la *voz* (sonido que produce el aire expelido de los pulmones al hacer vibrar las cuerdas vocales) objeto de estudio del presente proyecto, requiere que se agrupen los conceptos básicos relacionados con la anatomía del aparato fonador, su fisiología y la fonación, elementos que serán empleados para introducirnos en el tema del tratamiento de señales de voz por medio de la técnica *wavelet.*

1.1 LA VOZ HUMANA

La emisión de la voz es un fenómeno producto de la capacidad que poseen los órganos que conforman el aparato fonador del ser humano, órganos que comparten también funciones como la respiración y la deglución. La producción y emisión de los sonidos verbales se debe al funcionamiento de varios órganos secuenciados que sincrónicamente trabajan, donde una corriente de aire que proviene de los pulmones es transformada a su paso por el aparato fonador y que llega a convertirse en los sonidos adecuados para la comunicación humana.

Los mecanismos por los cuales se produce la voz son complejos, debido a ello es interesante estudiarlos desde el punto de vista anatómico y fisiológico, con la finalidad de entender mejor su funcionamiento, dando un conocimiento que facilite su comprensión.

1.2 EL APARATO VOCAL: ANATOMÍA[4] Y FISIOLOGÍA[5]

El aparato vocal es el conjunto de órganos que permiten la producción de la voz, organizados en 3 grupos diferenciados: Cavidades infraglóticas (órganos de respiración: pulmones, bronquios y tráquea); cavidades glóticas (órganos de fonación: laringe, cuerdas vocales y glotis) y cavidades supraglóticas (órganos de articulación: paladar, lengua, dientes y labios) [2]. Cada una de ellas realiza una función distinta e imprescindible en la fonación; a continuación se hará una breve descripción de cada una de ellas. En la figura 1.1, se puede ver la anatomía del aparato vocal.

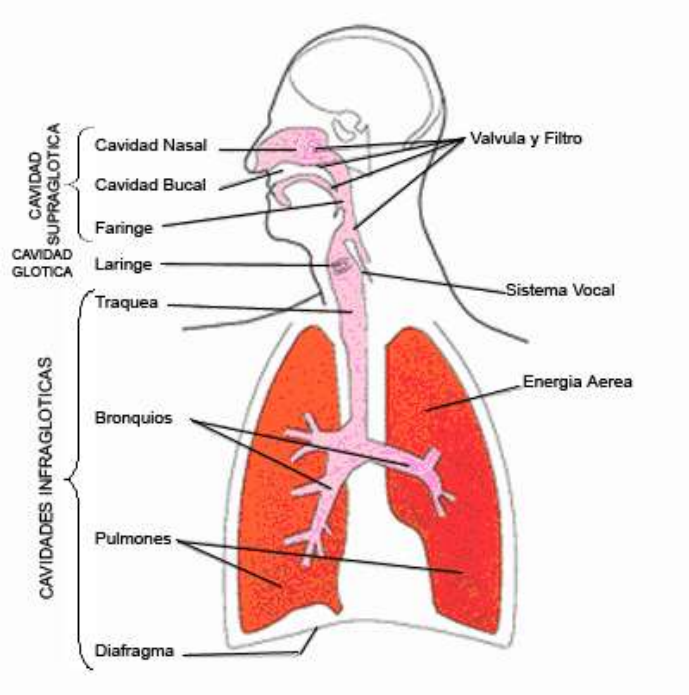


Figura 1.1 Esquema en conjunto del aparato vocal

[4] La anatomía es la ciencia que tiene por objeto dar a conocer la estructura, número, situación y relaciones de las diferentes partes del cuerpo orgánico. Real Academia de la Lengua Española, en *http://www.rae.es*.

[5] La fisiología es la ciencia que tiene por objeto el estudio de las funciones de los seres orgánicos. Real Academia de la Lengua Española, en *http://www.rae.es*.

1.2.1 Cavidades infraglóticas (Órganos respiratórios) [1][2]

Las cavidades infraglóticas son las cavidades situadas debajo de la glotis, que proporcionan la corriente de aire necesaria para producir el sonido, en éstas se encuentran los órganos de la respiración: diafragma, pulmones, bronquios y tráquea. El diafragma es un músculo situado por debajo de los pulmones y con forma de cúpula, su misión es controlar la expansión o despliegue de la cavidad pulmonar o su reducción y vaciado junto con la ayuda de los músculos pectorales y con ello la respiración. Cuando se contrae el diafragma se ensancha la cavidad torácica, produciéndose la inspiración de aire; al relajarse se reduce la cavidad, produciéndose la espiración del aire contenido en los pulmones.

Los pulmones juegan el papel más relevante, su misión es doble: por un lado, fisiológicamente son instrumento de la respiración con toda la serie de transformaciones bioquímicas que en ellos se originan; por otro lado, sirven de proveedores de la cantidad de aire suficiente para que el acto de la fonación sea realizable. El aire contenido en los pulmones va a parar a los bronquios y de aquí a la tráquea, los cuales son tubos cartilaginosos que conducen el aire entre los pulmones y la laringe; su función en la fonación es la de canales de conducción del flujo aéreo.

1.2.2 La cavidad glótica [1][2]

La cavidad glótica formada en la laringe, es el principal órgano de la voz [2], es una especie de caja cartilaginosa situada al final de la tráquea y que se conecta con la faringe, su función primordial es facilitar la obturación de la traquea; su movilidad le permite ascender o descender. La cavidad glótica está conformada por cartílagos unidos entre sí por ligamentos (tejido de haces fibrosos que mantiene a los órganos en la debida posición) y fascias[6], así como músculos recubiertos por una mucosa[7]. Los pliegues vocales forman parte de la laringe, denominados con frecuencia cuerdas vocales, que son propiamente dos músculos recubiertos por una mucosa, figura 1.2:

[6] Las fascias son láminas tendinosas que unen o envuelven los órganos internos, sobre todo los músculos [2].

[7] Una mucosa es el revestimiento que recubre un órgano hueco: estómago, traquea, faringe, boca, etc.[2] p15.

Figura 1.2 Esquema de la laringe

Ente los cartílagos de la laringe se encuentran:

- Cricoides: es la base, en forma de anillo.
- Tiroides (nuez o bocado de Adán), en forma de escudo.
- Dos aritenoides, que poseen gran movilidad.

Los pliegues vocales están unidos al tiroides y a los dos aritenoides, que se encargan de su movimiento. El espacio comprendido entre los pliegues vocales cuando están alejados uno de otro se denomina glotis. En la figura 1.3 podemos ver un corte transversal de la laringe.

Figura 1.3 Corte transversal de la laringe. Movimiento del cartílago aritenoides y de los pliegues vocales (líneas continuas o discontinuas)

El sonido es el efecto del trabajo conjunto del aire infraglótico y la tensión de las cuerdas, donde la vibración de los pliegues vocales se produce por la presión que ejerce el aire infraglótico sobre éstos pliegues que se encuentran están unidos, y que vibran para permitir el paso del aire a una determinada frecuencia (frecuencia fundamental FF), que es la misma frecuencia de la onda sonora que se origina; en este

sentido si las cuerdas vocales vibran producen sonidos sonoros y si no lo hacen producen sonidos sordos.

La frecuencia fundamental de la voz, es el parámetro más importante a tener en cuenta en el análisis de voz y habla [12], pues a partir de este es que se producen los sonidos que caracterizan los segmentos sonoros en la fonación. Cualquier perturbación en la frecuencia fundamental, se refleja inmediatamente en la salida de información y altera la correcta dicción.

1.2.3 Las cavidades supraglóticas [1][2]

Las cavidades supraglóticas corresponden a la faringe, la cavidad bucal y la cavidad nasal. La faringe está entre la parte posterior de la boca y la laringe, es una cavidad muscular capaz de estrecharse de atrás a adelante, lateralmente y también puede variar su volumen verticalmente dependiendo de los movimientos de elevación y descenso de la laringe que desempeñan una importante función en la articulación de las vocales,.

La cavidad bucal que está compuesta por la úvula, la lengua, velo del paladar, dientes y labios, donde la úvula, es el primer obstáculo para el aire proveniente de los pulmones. El velo del paladar al elevarse impide que el aire pase por la nariz; durante el habla permanece descendido para las vocales y las consonantes nasales (m, n y ñ) y se eleva para los demás sonidos.

La lengua es el órgano más móvil de la boca, registra una actividad elevada durante el habla y se divide en tres partes: raíz, dorso y ápice. Los dientes son órganos pasivos en la medida que están insertos en los maxilares; la mandíbula inferior, es un órgano activo dada su influencia en la articulación de los sonidos. Finalmente tenemos los labios, elementos que poseen bastante movilidad y que permite modificar los sonidos. Todo este conjunto de órganos, empezando por la laringe es lo que se conoce como el tracto vocal, que adoptará diferentes formas en función de las posiciones relativas de la mandíbula, la lengua, los labios y otras partes internas.

La cavidad nasal actúa como cavidad resonadora y esta compuesta por partes del cuerpo que vibran al estar en contacto con el sonido y cuyo tamaño y forma determina su frecuencia de vibración. Los sonidos producidos por el paso del sonido fundamental a través de las cavidades de resonancia (armónicos) enriquecen el timbre vocal. Las cavidades que intervienen en la resonancia son: La nasofaringe, fosas nasales, senos paranasales, maxilar superior, paladar y hueso frontal que refuerzan sonidos agudos; pabellón faringobucal que refuerza todos los sonidos en forma pareja; la faringe media y la región supraglótica que refuerza los sonidos graves.

En la figura 1.4 podemos apreciar la faringe, la cavidad bucal y la cavidad nasal, con sus partes.

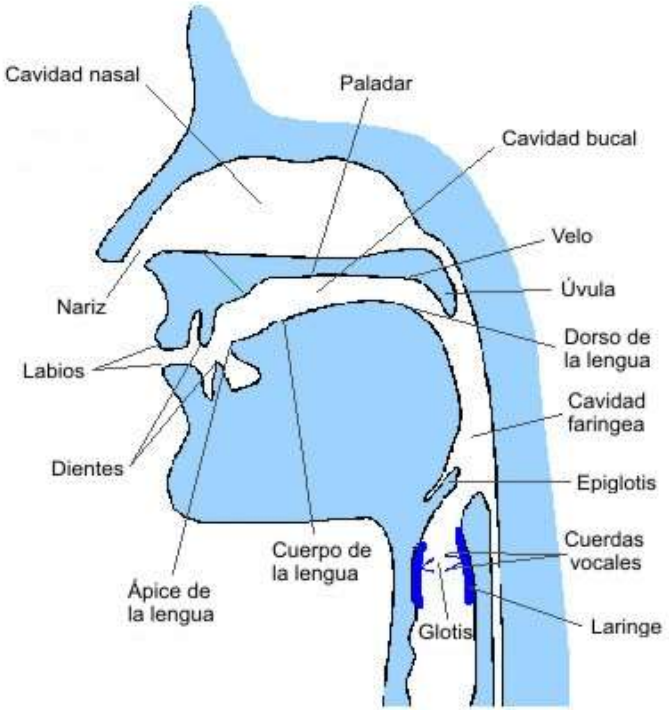


Figura 1.4 Faringe, la cavidad bucal y la cavidad nasal

En resumen y como se ha señalado, todo se inicia en los pulmones: El aire sale expulsado de ellos hacia la laringe a diferente presión en función del sonido que se pretende generar. La glotis separa los pliegues vocales y se mantiene abierta mientras respiramos, pero en el momento de producir sonidos se va estrechando, de manera intermitente, cerrando el paso del aire. Tras superar la glotis el aire se acerca al tracto vocal, que va variando su forma en función de cuales sean los sonidos que deseemos

producir. Los elementos articulatorios de la cavidad oral (lengua, labios, mandíbulas, velo del paladar) actúan como resonadores variables, que favorecen o neutralizan componentes espectrales de la onda de presión que hasta aquí haya llegado.

1.3 PRODUCCIÓN ACÚSTICA DEL HABLA

La emisión de la voz es un fenómeno de enorme variación, excluyendo las diferencias que existen entre una y otra persona, en un mismo individuo la voz toma múltiples aspectos, de acuerdo a sus diversas circunstancias psicofisiológicas, que se caracterizan por la existencia entre una actitud psicológica y determinados hechos específicos correspondientes a la actitud física y a la mecánica del soplo fonatorio. Esta producción acústica del habla se basa especialmente en el fenómeno de la resonancia. El resonador debido a sus características físicas (dimensión, forma, grado de rigidez de las paredes) posee una frecuencia natural de resonancia.

Con el objeto de ser más específicos en el estudio de conceptos y análisis de la producción acústica del habla, se considerarán sólo fonemas emitidos de forma aislada. Para la producción de estos fonemas existen tres fuentes de energía sonora [1]:

1. La fuente glotal
2. La fuente fricativa
3. La fuente plosiva

La fuente glotal es generada por la modulación de una corriente de aire pulmonar por la vibración de los pliegues vocales, la cual origina una onda compleja periódica. El espectro de esta modulación es discreto y está constituido por líneas separadas de la frecuencia fundamental FF con intensidades que decaen a razón de aproximadamente 12 db/octava [1]. La figura 1.5 muestra en el dominio del tiempo una onda glotal compleja.

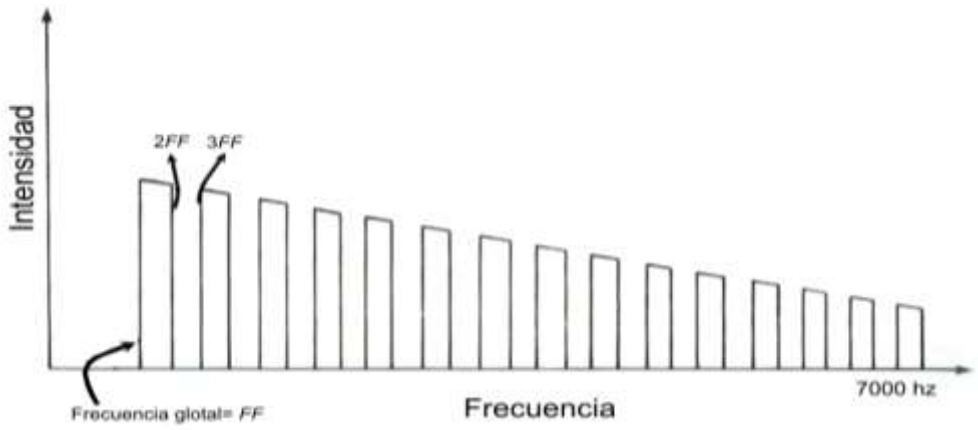


Figura 1.5 Onda glotal compleja en el dominio de la frecuencia

Se debe tener en cuenta que aunque la onda glotal parezca una onda sinusoidal, es en realidad una onda compleja. La figura 1.6 es el gráfico en el dominio del tiempo de la onda glotal.

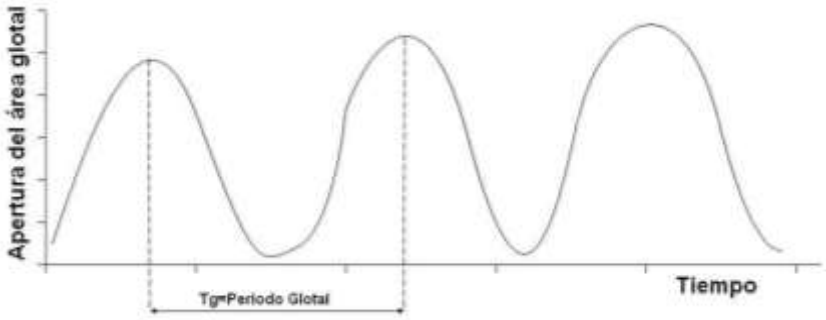


Figura 1.6 Onda glotal compleja en el dominio del tiempo

La fuente glotal por sí sola se utiliza para la producción de vocales y semivocales y combinada con la fuente fricativa y la plosiva se utiliza para la producción de consonantes sonoras de dichos tipos.

La fuente fricativa es generada por la turbulencia de la corriente de aire pulmonar forzada a través de una contracción en el tracto vocal. Durante esta producción fricativa la glotis se mantiene abierta. Esta fuente fricativa por sí sola se utiliza para la producción de consonantes sonoras tales como /s/ /f/ /y/ /z/.

La fuente plosiva es generada por la expulsión instantánea (explosión) de un caudal de aire retenido (implosión) a una presión mayor que la atmosférica debido a un cierre del tracto vocal. La fuente plosiva por sí sola se utiliza para la producción de consonantes, sordas tales como la /p/, /t/ y /k/. Su representación se muestra en la figura 1.7

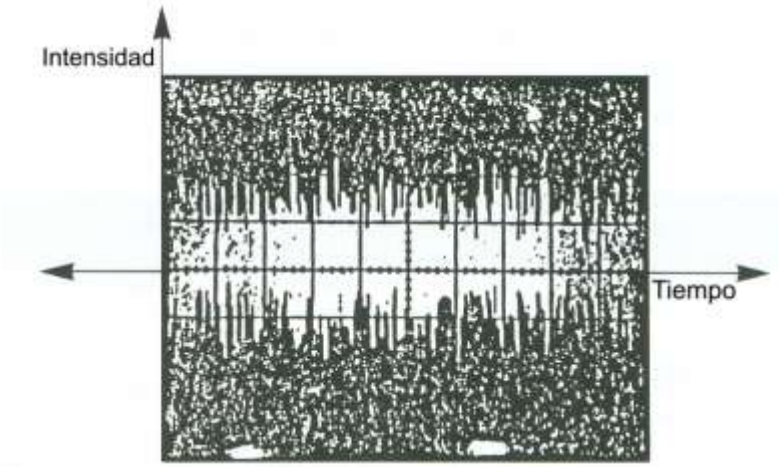


Figura 1.7 Señal generada por la fuente plosiva.

Estas fuentes de energía sonora por sí solas no servirán para la comunicación oral, de tal forma que sus espectros deben ser modificados a través de resonancia en las cavidades del tracto vocal y eventualmente nasal. Cada gesto articulatorio prepara dichas cavidades con una determinada curva de resonancia que amplifica y amortigua ciertas frecuencias de la fuente de energía sonora acopladas a ellas.

Para la estimación de la frecuencia fundamental en señales de voz fueron seleccionados los sonidos vocálicos, ya que según Jackson-Minaldi [1] y Cuene Hoover [12] éstos reúnen todas las características que se necesita para realizar la dicha estimación. Por otra parte, durante la articulación de palabras, la producción del tono fundamental se ve interrumpida, de ahí que los análisis del comportamiento del tono fundamental se realicen para segmentos sonoros con contenido invariable, con una vocal sostenida [12].

1.3.1 Sonidos vocálicos

Las vocales o sonidos vocálicos, son los fonemas sonoros del lenguaje, en los que el aire vibrante que sale por la glotis encuentra en la faringe, las fosas nasales y la cavidad bucal una caja de resonancia de dimensiones y forma variables para cada vocal, dichas señales fueron los sonidos escogidos para este trabajo por destacarse de otros sonidos del lenguaje. El sistema vocálico español conoce cinco fonemas: /i/, /e/, /a/, /o/, /u/ que presentan las siguientes características:

1. Las vocales son sonidos sonoros que están directamente relacionados con la vibración de los pliegues vocálicos, ya que la excitación está formada por un tren de pulsos cuasi-periódico generado por la vibración de las cuerdas vocales a diferencia de los sonidos no-sonoros (como las consonantes fricativas), donde la excitación es una señal aleatoria generada por el flujo turbulento del aire a través de las cuerdas vocales en relajación y los sonidos plosivos (como las consonantes /p/, /t/, etc.) donde la excitación se genera mediante una oclusión del tracto vocal seguida de un súbito aumento de presión y una relajación, lo que genera una excitación instantánea.
2. Su frecuencia fundamental se mantiene a pesar de los efectos de las cavidades resonadoras.
3. Son respuesta de la fuente de energía sonora glotal que por si sola se utiliza para la producción de las vocales y semivocales.
4. Estos fonemas son modificaciones exclusivas de la voz a diferencia de las consonantes que son sonidos que deben ir acompañados de otro en la cadena fónica, es decir, no siguen un criterio de producción sino de combinación.
5. Las vocales son fonemas sonoros y abiertos que se distinguen entre sí por su timbre[8] (onda sonora) característico.
6. Mientras que las consonantes establecen un obstáculo al paso del aire, las vocales se caracterizan por la ausencia de cualquier obstáculo.

[8] El timbre constituye la característica más relevante y distintiva del sonido y depende no sólo de la frecuencia de la onda acústica sino también de la naturaleza y forma del cuerpo que lo produce. En el caso de los sonidos articulados, el cuerpo que los produce es la boca o el aparato fonador en su conjunto.

1.4 TÉCNICAS DE ANÁLISIS.

Ya se ha definido el aparato vocal y establecido claramente que las señales de voz son de carácter *no estacionario*[9], por lo tanto ahora se realizará un reconocimiento de las técnicas clásicas con que se analizan estas señales, las cuales permiten la extracción de los parámetros acústicos que las caracterizan, estas técnicas son constantemente utilizadas para realizar el análisis de éstas señales, en el cual se convierten las señales de voz en una representación parametrizada de las mismas.

Dentro de las técnicas clásicas del tratamiento de señales se va a mencionar la técnica de análisis de la transformada de fourier en su forma discreta y mejorada, sin embargo existen otras técnicas de análisis que se basan en medidas tales como [1]: Cruces por cero, Autocorrelación, Análisis Cepstral, Banco de Filtros y Codificación por Predicción Lineal.

1.4.1 Transformada Discreta de Fourier (DFT)

La transformada discreta de Fourier (DFT, “Discrete Transform Fourier”) de *x(n)* se define como [19] [20] [40]:

$$X(k) = DFT\left\{x(n)\right\} = \sum_{k=0}^{N-1} x(n) e^{\left(\frac{-j2\pi kn}{N}\right)} \qquad (1.\ 1)$$

y su inversa (IDFT):

$$x(n) = DFT^{-1}\left\{X(k)\right\} = \sum_{k=0}^{N-1} X(k) e^{\left(\frac{j2\pi kn}{N}\right)} \qquad (1.\ 2)$$

con $k = 0,1,2,...,N-1$, donde N es el número de muestras de la ventana que se va a utilizar.

De esta manera los coeficientes de la transformada de Fourier son discretos y computacionalmente manejables, que pueden ser representados mediante una matriz de fourier por un vector que corresponde a valores discretos de una señal en el tiempo.

[9] Las señales no estacionarias, son señales cuya amplitud varía en forma rápida y abrupta en el tiempo o señales cuyo contenido de frecuencia es variable de un instante de tiempo a otro [4].

El motivo del uso de la DFT parte del hecho de la utilidad que tiene descomponer la señal de voz de partida en sus componentes en frecuencia.

Erell y Weintraub [21] describen un método para la utilización de la cuasi-periodicidad del habla, el algoritmo está estrechamente relacionado con la estimación del error cuadrático medio mínimo de la DFT, usado ya sea para mejorar el habla o reconocerla en presencia de ruido.

Como mejora al cálculo de la DFT, se encuentran algoritmos que la optimizan, como la transformada rápida de fourier *(FFT, Fast Fourier Transform,)* que realiza lo mismo que la transformada discreta pero en mucho menos tiempo [22], pretendiendo obtener una mejor resolución en el tiempo para señales semi-estacionarias o simplemente no estacionarias, lo cual constituye una poderosa herramienta para el análisis frecuencial de señales discretas, a la vez que elimina información redundante que genera la transformada discreta.

1.4.2 Técnica de análisis con cruces por cero

Es una de las técnicas más sencillas y consiste en contar la cantidad de veces que la señal cruza por el nivel de cero. La principal ventaja es el bajo costo computacional, pero cuando se trabaja con señales ruidosas o de bajo contenido frecuencial o señales donde alguno de los armónicos es más potente que la fundamental, sus resultados son poco precisos. Rabiner, Cheng, Rosenberg y Mcgonegal [6] notan además que las mediciones basadas en los cruces por cero pueden producir una pequeña variación o jitter, en el resultado cuando se presenta ruido en la señal de entrada.

Merlo, Fernández, Caram, Priegue y García [4] para el reconocimiento de voz de un individuo mediante una red neuronal de kohonen, utilizan el promedio de cruces por cero en el proceso de codificación de entrada para estimar la frecuencia fundamental y que posteriormente sirve para determinar el comienzo y fin de la palabra cuando la misma empieza o termina con sonidos de baja energía y alto valor de frecuencia, como lo son las fricativas 'f', 's', 'y', 'z', aunque se afirma que es una técnica algo rudimentaria. En el desarrollo de implantes cocleares para pacientes con sordera severas [3], se

utiliza esta técnica para obtener la frecuencia fundamental de la señal de voz mediante un filtro pasa baja de 270Hz, y la frecuencia de la segunda formante mediante un filtro pasa banda de 1Khz – 44Khz, de estos resultados se proporcionan estímulos en varios nervios referentes al oído para el proceso de escucha del paciente.

Esta técnica también es usada en combinación con otras para determinar la existencia de tramas de voz y/o silencios como lo señala Zanuy [7], donde exponen 2 algoritmos: el primero de Rabiner y Cambur, usado para el reconocimiento de palabras aisladas, y el segundo es una herramienta de compresión de silencios donde el módulo detector de actividad de voz realiza una clasificación como trama de voz activa o trama de voz no activa en función de la variación de cuatro características de la señal de entrada: La energía de la banda, la energía de las bajas frecuencias, los coeficientes LSP (Pares de Líneas Espectrales) y la tasa de cruces por cero.

1.4.3 Autocorrelación

Para una señal deterministica la autocorrelación se define como [8] [9] [40]

$$R_x[k] = \sum_{m=-\infty}^{\infty} x[m]x[m+k], \qquad (1.\ 3)$$

para una señal aleatoria se define como

$$R_x[k] = E\left[x[m]x[m+k]\right], \qquad (1.\ 4)$$

donde $R(0)$ es igual a la energía (en señales deterministicas) o igual a la potencia media (en señales periódicas o aleatorias).

Para definir la existencia de tramas de sonido o silencio se utiliza la relación entre los dos primeros valores de la autocorrelación de la señal: Se toma la relación $R(0)/R(1)$ como indicativo de la periodicidad de la señal [8], donde este parámetro da una medida de “blancura" de la señal y asociando segmentos sordos con segmentos semejantes a ruido se puede extraer información de la sonoridad. Ortega y González [11] se refiere a la estimación de la FF con autocorrelación, acentuando los máximos temporales de la señal (correspondientes al período fundamental), de forma que se pueda diferenciar con mayor claridad la FF del tramo analizado.

La autocorrelación se emplea para detectar la frecuencia fundamental en tramos sonoros, sin embargo, su información acerca de la envolvente espectral (tracto vocal), es más de la necesaria, ya que por ejemplo, el primer formante puede interferir con la estimación de la FF [10]**.** Es así como se hace necesario realizar un preproceso de la señal para eliminar influencias del tracto vocal y así obtener una detección del período fundamental mucho más clara; éste preproceso es llamado “Center Clipping” y consiste en recortar la señal entre cierto margen para reducir la información pero sin eliminar los picos necesarios para detectar los periodos; el inconveniente de este método es que puede realzar algunos picos de ruido, provocando errores de detección. [10] [8]

1.4.4 Análisis Cepstral

El análisis Cepstral se fundamenta en la suposición de que la voz es el resultado de la convolución de una función de excitación (que es generada en los pulmones) con la respuesta impulsional del tracto vocal. De este modo se quiere deconvolucionar las señales de voz para obtener por una parte la señal de excitación y por otra la respuesta del tracto vocal.

El análisis Cepstral es utilizado para la determinación de la FF. El Cepstrum se obtiene aplicando primero un enventanamiento, después se usa la Transformada de Fourier Discreta (DFT) para la señal, seguidamente se aplica el logaritmo a su espectro de potencia y por último se lleva al dominio del tiempo en un proceso inverso a través de la Transformada Discreta de Fourier Inversa (IDTF)[16]. El procedimiento se muestra en la figura 1.7

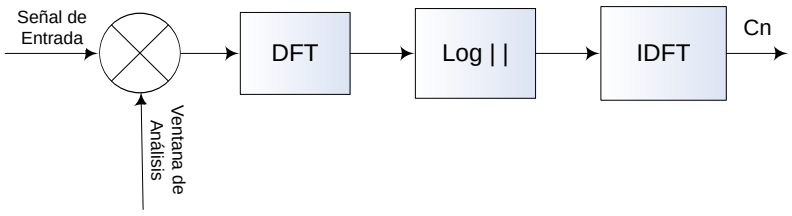


Figura 1.8 Análisis Cepstral

Utilizando una versión modificada de algoritmos basados en cepstrum, Ahmadi y Spanias [14] proponen un algoritmo de detección de la FF en dos pasos, primero se escoge un trozo de señal mediante una ventana móvil, luego se realiza un promedio de

estimación de umbral mediante tres procesos en paralelo: estimación de umbral de cruce por cero, estimación de umbral de cepstrum y estimación de umbral de Energía. En la segunda etapa de este algoritmo, se calcula el cepstrum de una ventana móvil, posteriormente se aplica un bloque de elección de picos de los valores de cepstrum y luego se realiza un procesamiento de los valores escogidos, teniendo en cuenta los tres umbrales calculados en la primera etapa y finalmente los valores son promediados. La desventaja de este algoritmo es que se requiere de una gran cantidad de procesamiento de la señal, dada las dos etapas del algoritmo en el sistema.

San martín y Carrillo [15] presentan un sistema de reconocimiento de palabras aisladas dependiente del locutor, las palabras se codifican mediante las técnicas de cepstrum y Predicción Lineal. De la técnica cepstrum se obtienen las bajas componentes que corresponden a variaciones lentas de las componentes espectrales y en consecuencia contienen información de la *envolvente del espectro*, la cual se relaciona con la respuesta en frecuencia del filtro que modela el tracto vocal, permitiendo reconocer características de dicho tracto vocal y los patrones de voz.

Faúndez y Fernandez [13] realizan un estudio de la influencia del ruido y de la variación temporal en reconocimiento de locutor, donde se usan diversas parametrizaciones de la señal de voz. Con el fin de obtener una parametrización robusta, independiente a las variaciones del entorno de grabación como el ruido, se utilizan coeficientes cepstrum, justificando su uso en la aplicación del logaritmo en el dominio espectral, donde simulando la respuesta aproximada logarítmica del oído respecto a la intensidad, aportan mejoras substanciales de las tasas de reconocimiento del locutor.

1.4.5 Banco de filtros

Los bancos de filtros son arreglos de filtros pasa bajas, pasa bandas y pasa altas para la descomposición espectral y composición de señales. Ellos juegan un papel importante en muchas aplicaciones modernas de procesamiento de señales tales como codificación de audio e imágenes y han sido históricamente la primera aproximación al procesamiento del habla; un banco de filtros pasa banda puede entenderse como un modelo sencillo de las etapas iniciales del sistema auditivo humano donde la señal

inicial se descompone en un conjunto discreto de muestras espectrales, que contienen una información similar a la que se presenta en los niveles superiores del sistema auditivo.

Esta técnica es muy conocida debido al hecho que permite fácilmente la extracción de componentes espectrales de una señal a la vez que provee eficientes implementaciones [24]. Uno de los algoritmos de codificación más conocidos para la compresión de audio es el MPEG audio, el cual utiliza un sistema de banco de filtros de 32 bandas en conjunto con la FFT como analizador de espectro para calcular la curva de enmascaramiento que se utiliza como umbral y que se basa en la percepción auditiva del oído humano, dejando pasar sólo las componentes de frecuencia dominantes [16]. Shiu, Yeh y Kuo [5] extraen con un banco de filtros de ocho bandas la frecuencia de cruces por cero que son usados para caracterizar el contenido de audio a través de las frecuencias dominantes en cada subbanda. Kaschel, Watkins y San Juan [18] en el trabajo de compresión de voz mediante técnicas digitales para el procesamiento de señales y aplicación de formatos de compresión de imágenes, es fundamental la aplicación de bancos de filtros ya que se consigue quitar redundancia a la señal de voz permitiendo mayor comprensión de la información al convertir una trama de voz filtrada en una imagen con el formato de compresión de imágenes que se seleccionó.

1.4.6 Codificación por predicción lineal (LPC)

La codificación por predicción lineal permite obtener información sobre la estructura acústica de los sonidos del habla, formando una aproximación razonable de ésta; se representa directamente en términos de parámetros relacionados con la función de transferencia del tracto vocal y las características de la función de la fuente que varían con el tiempo. Con un número suficiente de parámetros el modelo de predicción lineal puede constituir una aproximación adecuada a la estructura espectral de todo tipo de sonidos.

San Martín y Carrillo [15] presentan un sistema de reconocimiento de palabras aisladas dependiente del locutor. Cada palabra se codifica mediante las técnicas de Predicción Lineal y Cepstrum real y son los coeficientes de predicción los que se usan como

parámetros de reconocimiento de palabras. Parte del estudio de Rufiner y Milone [23] consistió en obtener patrones formánticos para ser utilizados como normativa en estudios de voces normales y patológicas, de los cuales utilizaron diversas técnicas para el análisis de las formantes y donde el método de LPC fue utilizado en la medición de los contornos formánticos.

2. TÉCNICA *WAVELET* EN EL PROCESAMIENTO DE SEÑALES DE VOZ

Existen muchos tipos de señales susceptibles de ser procesadas para extraerles información como son por ejemplo las señales sísmicas, de radar, bursátiles, satelitales, acústicas, electrocardiográficas (ECG), electroencefalográficas (EEG), imágenes (señales bidimensionales), etc. Generalmente estas señales presentan una amplitud que cambia de forma rápida y abrupta en el tiempo y/o cuyo contenido de frecuencia varia de un instante a otro, lo que las define como señales no estacionarias [25].

Teniendo en cuenta que cualquier señal puede ser representada por versiones escaladas y trasladadas de una *wavelet* madre, en este capítulo, se expondrá la teoría necesaria para *el análisis de las señales de voz,* con la finalidad de utilizar la técnica *wavelet* para la caracterización de la frecuencia fundamental.

2.1. *WAVELETS*

El término ***wavelet*** significa onda pequeña cuya función de carácter oscilatorio es de longitud finita. En este sentido, Subhasis Saha nos da una definición bastante sencilla: "Las *wavelets* son funciones definidas en intervalos finitos que tienen un valor promedio de cero" [26], es decir:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{2.1}$$

Sin embargo, para que una función sea posible candidata a ser *wavelet* (ψ) debe cumplir ciertas condiciones: Que la función ψ sea continua, con momentos nulos, que decrezca rápidamente hacia cero cuando la variable independiente tienda hacia infinito, o que sea nula fuera de un segmento de $\mathbf{R}$. Más precisamente, la condición de admisibilidad para que la función ψ pertenezca al espacio de las funciones de energía finita es:

$$\int_{-\infty}^{0} \frac{\left|\hat{\psi}(s)\right|^2}{|s|} ds = \int_{0}^{+\infty} \frac{\left|\hat{\psi}(s)\right|^2}{|s|} ds = K_{\psi} < +\infty \quad . \tag{2.2}$$

Para conocer la forma de la *wavelet* y su habilidad para suprimir un polinomio dado, se necesita reconocer una característica matemática muy importante de la *wavelet* llamada ***momentos de desvanecimiento***, útil para propósitos de compresión, donde la suavidad de la *wavelet* esta limitada por el número de desvanecimientos. Inicialmente es suficiente pensar en el "momento" como una extensión del "promedio"; lo cual quiere decir que después de que el valor promedio de la *wavelet* sea cero, tiene (al menos) un momento de desvanecimiento. El i-ésimo momento de la *wavelet* se calcula con la siguiente integral:

$$\int_{-\infty}^{+\infty} t^i \psi(t) dt = 0 \quad . \tag{2.3}$$

Es decir, si el valor promedio de $t^i \psi(t)$ es cero (donde $\psi(t)$ es la función *wavelet*), para $i = 0,...,n,$ entonces la *wavelet* tiene $n+1$ momentos de desvanecimiento y los polinomios de grado n son suprimidos por esta función, lo cual indica que todas las señales que tengan la forma polinomial del tipo:

$$f(t) = \sum_{m=0}^{n+1} c_m t^m \tag{2.4}$$

tienen cero coeficientes *wavelet* [25].

La ***regularidad*** es otra característica de las *wavelet*s. En estudios prácticos y teóricos, la noción de regularidad ha ido incrementando en importancia dado que es útil para obtener características nítidas como la suavidad en las señales o imágenes reconstruidas y para la función estimada en análisis de regresión no lineal[10]

[10] El objetivo de la regresión es obtener un modelo de la relación entre una variable **Y** o más variables **X**. El caso más simple es la regresión lineal f(X)=Y=aX+b+e, donde f es la función que asigna los valores. Cuando f es totalmente desconocida, el problema es de regresión no lineal y la relación se convierte en un problema no paramétrico y puede ser resuelto: usando las técnicas usuales de enventanamiento estadístico o mediante los métodos basados en *wavelet*s

Para hallar la regularidad (R) de una función ψ en una zona determinada por los puntos (a,b), se suman las derivadas de la función en los puntos $(a_1,a_2,a_3,...b)$ que están entre a y b y se divide sobre el producto del número de puntos por la derivada del último punto tomado, es decir, la derivada de ψ en b, ya que así se podrá ver cómo varía la pendiente de la tangente a la curva de esos puntos:

$$R_{a,b} = \frac{\psi'(a_1)+\psi'(a_2)+\psi'(a_3)+...+\psi'(b)}{n.\psi'(b)} = \frac{\sum_{i=a}^{b}\psi'(i)}{n.\psi'(b)} \quad , \qquad (2.5)$$

donde $\psi'(b) > \psi'(a)$, siendo $\psi'(a_1)$ y $\psi'(b)$ las derivadas en los puntos a y b, con a_2, a_3 las derivadas en los puntos intermedios de (a,b) y con n el número de derivadas que se suman; pudiéndose tomar cualquier número de derivadas entre a y b. Entre más puntos se tomen entre a y b mayor será la precisión.

Para que una función sea regulable debe cumplir con las siguientes características:

- La función tiene que ser derivable en los puntos donde se elige observar la regularidad.
- Las funciones constantes no son regulables ya que la derivada mayor será cero y se dará una indeterminación del tipo: $\frac{0}{0}$.
- De la anterior característica se deduce: Que las funciones son mas regulares en ciertos puntos que en otros (ver Figura 2.1); es decir, no se puede realizar la regularidad donde la función tenga pendiente cero; pero sí en otras zonas.

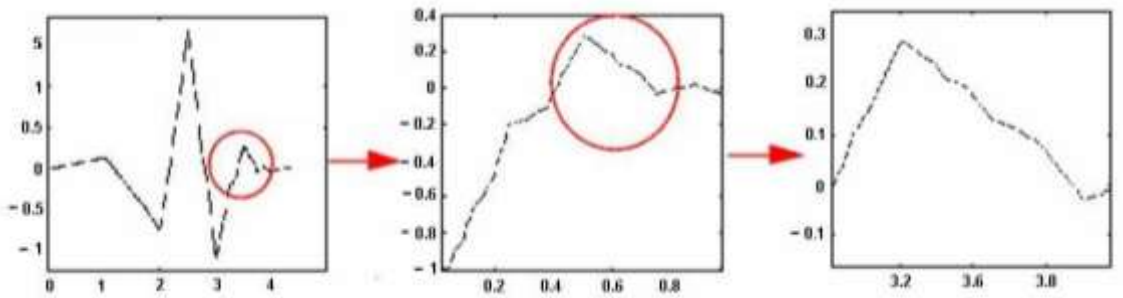


Figura 2.1 Zoom en la *wavelet* Daubechies3.

La regularidad de ciertas *wavelet*s es conocida, por ejemplo en la Tabla 2.1 se muestran algunas indicaciones de la regularidad de las *wavelet*s Daubechies.

***wavelet* (ψ):**	db1=Haar	db2	db3	db4	db5	db7	db10
Regularidad:	Discontinua	0.5	0.91	1.27	1.59	2.15	2.90

Tabla 2.1 Indicaciones de la regularidad de las *wavelet*s Daubechies.

Entre otras propiedades de las *wavelet*s se destacan las siguientes:

- ✓ La propiedad de ***soporte compacto***, la cual define que la *wavelet* sea de duración finita, es decir, la velocidad de convergencia a cero de $\psi(t)$ o $\hat{\psi}(w)$ cuando el tiempo t o la frecuencia w tienden a infinito. Esta determina la cantidad tanto de localizaciones en tiempo y frecuencia permitiendo una menor complejidad en los cálculos y una mejor resolución en tiempo y frecuencia.
- ✓ La propiedad de ***simetría*** que permite que los filtros sean de fase lineal, útil para evitar el desfasamiento en el procesamiento de las señales.
- ✓ La propiedad de ***ortogonalidad*** que es la característica que se logra cuando el producto punto de dos vectores es igual a cero. Esta es importante en este tipo de estudios para que los análisis sean estables.
- ✓ La propiedad de ***ortonormalidad*** relacionada con las *bases ortonormales* que son un conjunto de vectores ortogonales cuya norma[11] es igual a 1; por ejemplo: <(1,0,0);(0,1,0);(0,0,1)>. Es posible hallar una base ortonormal a partir de una base ortogonal dividiendo a cada vector de la base ortogonal por su norma.

Las *wavelet*s que cumplen con las anteriores características, se definen por medio de una o varias funciones iniciales, llamadas *wavelets* madre (ψ); al hacer referencia al término 'madre', se está indicando el hecho de que las funciones usadas derivan de una función principal, es decir, la *wavelet* madre es un prototipo a partir del cual se generan otras funciones con características similares [27][28].

En la ecuación 2.6, la expresión $\psi_{a,b}(t)$ es un prototipo de función *wavelet* en escala a y con un desplazamiento espacial b :

[11] Un vector es un elemento de un espacio vectorial para el que, en ocasiones, interesa conocer su longitud. Esto es lo que hace el **operador norma**: determina la longitud del vector bajo consideración.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left[\frac{t-b}{a}\right] \tag{2.6}$$

Gracias al reconocido buen desempeño de la técnica wavelet en comparación con otras herramientas de análisis como la Técnica de Fourier, esta técnica puede utilizarse para limpieza de ruido, análisis y compresión señales, etc, incluso en un entorno donde dichas señales sean esporádicas, es decir, no periódicas. Esto se puede hacer siempre y cuando la *wavelet* sea lo más parecida (tenga un buen nivel de correlación) al tipo de señal que se desea analizar.

2.2 TRANSFORMADA *WAVELET*

El objetivo de la transformada *wavelet* es descomponer cierta señal en señales componentes denominadas *wavelets,* las cuales forman una base del espacio de funciones, con ciertas propiedades como ortogonalidad, tamaño, suavidad, duración, etc. Así, este método sería una ampliación del método de *Fourier,* en el que la señal se descompone en funciones senoidales. En este caso, la descomposición se realiza a partir de funciones más complejas, en las cuales además no se varía su frecuencia, sino su posición y su escala temporal [9]. La transformada *wavelet* se clasifica en continua y discreta.

2.2.1 Transformada *wavelet* continua

La transformada *wavelet continua* (TWC) se usa en el análisis de señales que buscan obtener un conjunto más o menos reducido de coeficientes $C(a,b)$ que caractericen adecuadamente las señales $f(t)$ a través de la *wavelet* madre seleccionada.

$\psi_{a,b}(t)$ tiene un prototipo genérico de función *wavelet* en escala 'a' y con un desplazamiento espacial 'b':

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left[\frac{t-b}{a}\right] \tag{2.7}$$

Ahora, la transformada *wavelet continua* esta definida de la siguiente manera [20]:

$$C_{a,b}(t) = \int_{-\infty}^{+\infty} f(t)\frac{1}{\sqrt{a}}\psi\left[\frac{t-b}{a}\right]dt \tag{2.8}$$

La idea básica de esta transformada es que mediante las dilataciones (variaciones del parámetro 'a') y las traslaciones (variaciones del parámetro 'b') en el tiempo de una función *wavelet* madre (ver Figura 2.2) se pueda representar cualquier función, es decir, a partir la *wavelet* madre es posible generar el resto de funciones de la familia mediante cambios de escala y traslaciones logrando que estas se parezcan lo más posible a las señales que se desean analizar. La TWC se presenta como una herramienta de análisis de señales con capacidad de localización *tiempo-frecuencia* variable comparada con la localización fija de la transformada de Fourier ventaneada.

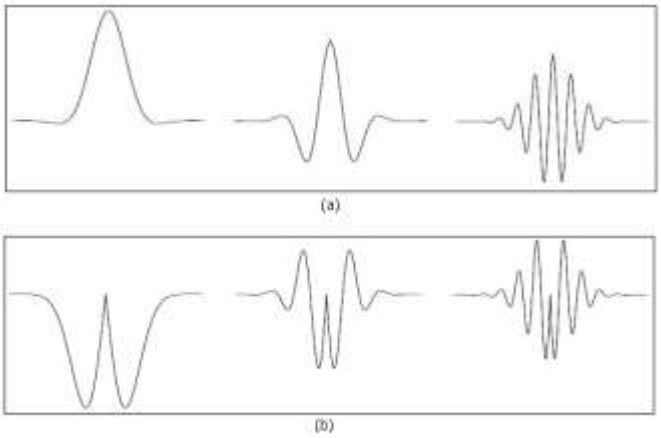


Figura 2.2 Traslación y dilatación de una función *wavelet* madre compleja.
(a) Parte Real, (b) Parte Imaginaria.

2.2.2 Transformada *wavelet* Discreta

El Procesamiento Digital de Señales esta definido como el conjunto de técnicas y herramientas para el tratamiento de señales en el dominio discreto o digital [28], en nuestro caso la herramienta para el procesamiento de señales de voz será el computador.

Debido a que los computadores trabajan sólo con datos discretos, el cálculo numérico de la transformada w*avelet* de la señal de voz requiere valores discretos o muestras, en consecuencia, la transformada arrojará solo valores discretos de la señal. Además, en términos de cálculo computacional es imprescindible discretizar la transformada y la suposición más lógica es que tanto los valores de la escala como los de la traslación sean discretos, significando esto que se va a tener la transformada *wavelet* en su versión discreta TWD. Dado que desea llevar a cabo un recubrimiento discreto del plano tiempo-frecuencia y que el recubrimiento localizado sea diferente a cada escala, la discretización del parámetro de traslación b dependerá del parámetro de escala a. Para escalas mayores, la traslación deberá ser mayor (ver Figura 2.3).

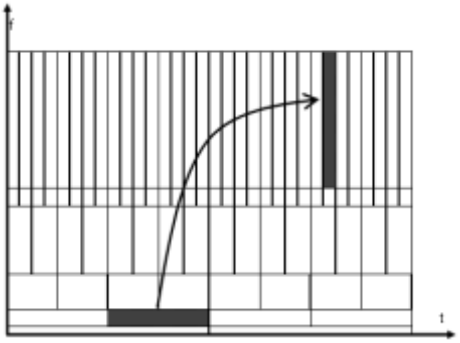


Figura 2.3 Recubrimiento en el plano *wavelet.*

La forma más común de discretizar los valores de a y b es utilizando una red diádica[12], es decir, $a = 2^{-j}$ y $b = k2^{-j}$ con j y $k \in \mathbf{Z}$, de tal manera que el conjunto de funciones de la ecuación 2.7 se transforma en:

$$\psi_{j,k}(n) = 2^{\frac{j}{2}} \psi\left(2^j n - k\right) \tag{2.9}$$

Las cuales corresponden a la versión de la función *wavelet* diádicamente discretizada.

2.3 FAMILIA *WAVELET* PARA EL TRATAMIENTO DE LAS SEÑALES DE VOZ

Para el tratamiento de señales en este trabajo se eligió primero el tipo de señal de voz a analizar y segundo la *wavelet madre* que formó la familia *wavelet* para el análisis de

[12] La red diádica es cuando el factor de escala a es igual a 2 y partir de esta escala de referencia, todas las escalas que se utilizan en el sistema son potencias negativas de 2 [8].

dichas señales, en el primer caso se seleccionaron las vocales porque son las señales de voz más claras y sonoras y fisiológicamente representan mucho mejor la vibración glotal que es lo que determina la frecuencia fundamental y cuyo fundamento funcional y orgánico se sustentó en el capitulo 1. Los argumentos para la selección de la *wavelet* madre se presentan a continuación.

2.3.1 Selección de la *wavelet* madre

Como mencionó anteriormente, dependiendo del escenario donde se quiera aplicar la transformada *wavelet* existen diferentes tipos de señales que se pueden analizar y dependiendo de ellas se escoge la *wavelet madre* que más se adapte a las características de esas señales. Debido a que la confiabilidad del método de análisis *wavelet* depende de la función *wavelet* madre utilizada, es necesario escoger la *wavelet* mas adecuada para el procesamiento y análisis de las señales de voz.

Leonard J. García en su trabajo “Transformada *wavelet* aplicada a la extracción de información en señales de voz” [8], identifica aspectos que caracterizan la voz como *la frecuencia fundamental* utilizando la *wavelet Gaussiana compleja*, enfocado su interés en la *wavelet* madre que permita obtener un conjunto más o menos reducido de coeficientes que caractericen adecuadamente las señales, gracias a la comparación que hace de los resultados con los resultados de otros sistemas de parametrización buscando no superar la veintena de coeficientes resultantes de la descomposición. La decisión, de optar por ésta *wavelet* se justifica en que es la función que mejor mantiene la resolución en los dominios de tiempo y frecuencia; y también porque otros trabajos presentan soluciones clásicas basadas en dicha función como lo son la función Sombrero Mejicano o función de Morlet [8].

Juan G. Jaramillo y Gustavo A. García en su proyecto “Reconocimiento de hablantes usando transformada wavelet y DSP’S” [29] tienen en cuenta dos criterios para la selección de la *wavelet* madre: primero la mayor similitud en las muestras de señales de voz y segundo la menor carga computacional lo cual hace referencia a la longitud de los filtros. Con respecto al primer criterio se descartaron la *wavelet* Haar por su gran diferencia entre las señales de voz y la *wavelet* Symlet por su simetría ya que las

señales de voz son asimétricas, planteando como una buena opción la *wavelet Daubechies* y la *Coiflet*. Con respecto al segundo criterio la longitud de los filtros de las *wavelet*s *Daubechies* y las *Symlets* es dos veces el orden de las *wavelets* (2N) y para la *wavelet Coiflet* la longitud de sus filtros es de tres veces su orden (3N); teniendo en cuenta que el orden alto de la función *wavelet* implica una mayor longitud en los coeficientes de los filtros y por tanto un mayor costo computacional, por ello fue seleccionada la *wavelet* Daubechies 8 para estimar la frecuencia fundamental.

Alexander F. Sepúlveda y Germán Castellanos consideraron en su trabajo titulado "Estimación de la frecuencia fundamental de las señales de voz usando transformada *wavelet*" [30], la relación que existe entre una menor longitud del soporte compacto contra un mayor número de momentos de desvanecimiento; destacando la *wavelet Daubechies* y la de tipo *Spline* como las que mejor se desempeñan en este sentido.

Dentro de la familia de las *Spline* se selecciona, en calidad de la mejor *wavelet* madre, aquella *wavelet* que entregue la mayor cantidad de coeficientes cercanos a cero, de tal forma que solo unos pocos sean de un alto valor respecto de los demás. Con este propósito los autores usan una medida de variabilidad de la energía. Es así, como se calcula la menor función de costo final, que es el resultado de aplicar la función de entropía de Shannon[13] en cada escala de descomposición y la suma total de todos los valores por todas la escalas de descomposición, para seleccionar la *wavelet* cuyo orden dé el menor valor en la función de costo. Esto es:

$$C = \min_{k} \sum_{\lambda=1}^{J} C_{k,\lambda} \tag{2.10}$$

De donde k corresponde con la *k-ésima wavelet* madre a probar y $C_{k,\lambda}$ está dada por:

$$C_{k,\lambda} = -\sum_{m=1}^{N} \frac{\left|\left\langle f, \psi_{m,\lambda} \right\rangle\right|^2}{\left\|B^{\lambda}\right\|^2} \log_e \frac{\left|\left\langle f, \psi_{m,\lambda} \right\rangle\right|^2}{\left\|B^{\lambda}\right\|^2}, \tag{2.11}$$

donde f es la señal de voz, $\left\langle f, \psi_{m,\lambda} \right\rangle$ son los coeficientes wavelets;

[13] En teoría de la información, es una medida de la información contenida en un mensaje.

J. C. Long y S. Datta en el trabajo "*Wavelet* based feature extraction for phoneme recognition" [31] realizan la extracción de la *frecuencia fundamental* modelando los fonemas del habla mediante la *wavelet* de *Morlet*. Para este proceso se tuvo en cuenta que la señal fuera suave y semiperiódica.

La exposición de las anteriores referencias, las funciones *wavelets* citadas en el anexo B y toda la consulta bibliográfica realizada sobre la temática nos permitieron seleccionar para este trabajo la *wavelet gaussiana compleja*, gracias a que reconocimos en ella que reúne la mayoría de las características mencionadas: Esta función se consideró la más apropiada ya que permite una buena resolución tanto en tiempo como en frecuencia y tiene similitud con las señales de voz. Se descartaron las wavelets daubechies, spline y morlet para estimar la FF, debido a que experimentalmente se estableció que no eran tan precisas como la gaussiana compleja.

2.3.2 *Wavelet* Gaussiana Compleja

Esta *wavelet* esta dada por la función densidad de probabilidad de Gauss o función gaussiana, expresada en la siguiente ecuación:

$$f\left(t;\sigma,\mu\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(-\frac{\left(t-\mu\right)^2}{2\sigma^2}\right)}, \qquad (2.12)$$

donde σ es la varianza, μ es la desviación estándar. Genéricamente esta función esta definida así:

$$f\left(t\right)=Ce^{-\frac{t^2}{2}}, \qquad (2.13)$$

donde C es un coeficiente de normalización. Esta función es regular y simétrica, no posee soporte compacto y no es ortogonal, como se muestra en la figura 2.4.

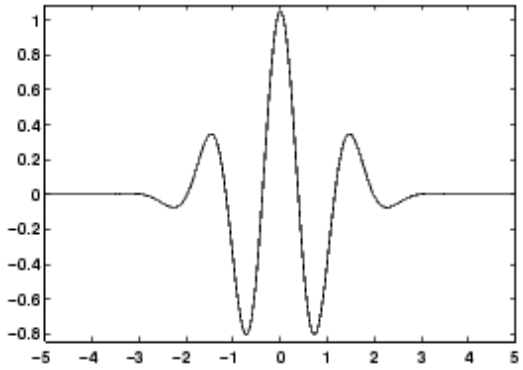


Figura 2.4 Función Gaussiana.

A partir de la ecuación (2.13) y teniendo en cuenta que la función compleja tiene un componente imaginario, la wavelet compleja gaussiana se define así:

$$f(t) = C_p e^{-jt} e^{-t^2} \quad , \tag{2.14}$$

donde C_p es tal que:

$$\left\| f^{(p)} \right\|^2 = 1 \quad , \tag{2.15}$$

donde $f^{(p)}$ es la derivada de orden p de f. En la figura 2.5 se aprecia la parte real (a) y la parte imaginaria (b) de la wavelet gaussiana compleja.

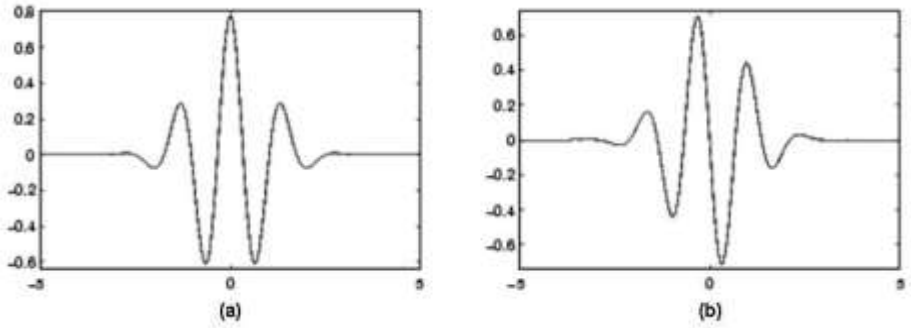


Figura 2.5 *wavelet* compleja gaussiana de orden 8. (a) Parte real. (b) Parte Imaginaria.

2.4 EL OÍDO HUMANO COMO MODELO DE ANÁLISIS

Para el procesamiento de las señales de voz llevado a cabo en este trabajo se tuvo como modelo al mejor arquetipo para la percepción de frecuencias, el aparato de audición humana, donde tenemos en cuenta la fisonomía y fisiología del oído humano

para implementar el sistema *wavelet* para la caracterización de la frecuencia fundamental de las señales de voz.

Figura 2.6 Partes del sistema auditivo humano.

El sistema auditivo es uno de los cinco sentidos del sistema sensorial, el cual tiene la capacidad de percibir el sonido y consta de tres partes: oído externo, oído medio y oído interno (ver figura 2.6). En este último se encuentra la cóclea (también conocida como caracol) es una estructura en forma de tubo enrollado en espiral que contiene el órgano de Corti, que esta conformado por alrededor de 24.000 a 30.000 células ciliadas que descansan sobre la membrana basilar (ver figura 2.7).

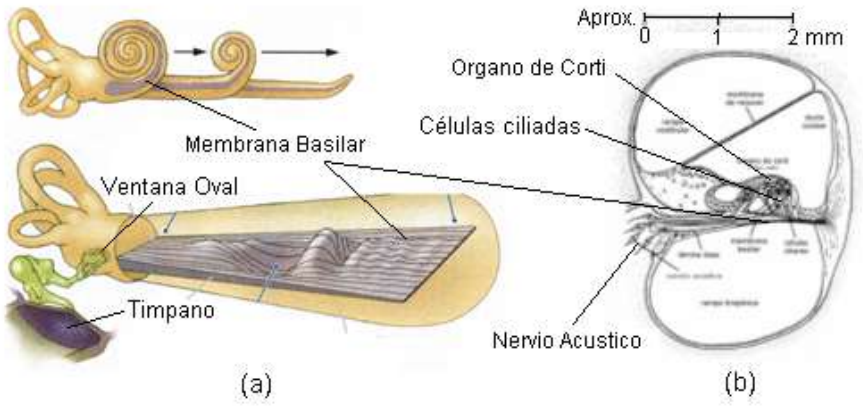


Figura 2.7 Estiramiento de la Cóclea: (a) Corte longitudinal y (b) Corte transversal.

La membrana basilar es la responsable del análisis en frecuencia de los sonidos que llegan al oído medio, sin embargo, las células capaces de decodificar esta información y enviarla al cerebro, se hallan ubicadas en el órgano de Corti. Esta respuesta es posible gracias a que la membrana varía en masa y rigidez a lo largo de toda su longitud, con lo que su frecuencia de resonancia no es la misma en todos los puntos: la membrana es rígida y ligera en el extremo más próximo a la ventana oval y al tímpano, por lo que su

frecuencia de resonancia es alta y en el extremo más distante la membrana es suave y pesada por lo que la resonancia es de baja frecuencia (ver figura 2.8).

Figura 2.8 Respuesta a frecuencias de la membrana Basilar.

El margen de frecuencias de resonancia disponible en la membrana basilar determina la respuesta en frecuencia del oído humano, las audiofrecuencias van de 20 a 20.000 Hz en este margen, la zona de mayor sensibilidad del oído se encuentra entre 1.000 y 5.000 Hz, teniendo mayor sensibilidad a los tonos agudos. La respuesta en frecuencia del oído permite la capacidad de tolerar un rango dinámico que va desde 0 dB (umbral de audición) a 120 dB (umbral de dolor).

2.4.1 Escala de Bark

La escala de Bark es un sistema de medición psicoacústica[14] que maneja un rango de 1 a 24 bandas correspondientes a las primeras 24 bandas críticas del oído (ver figura 2.9).

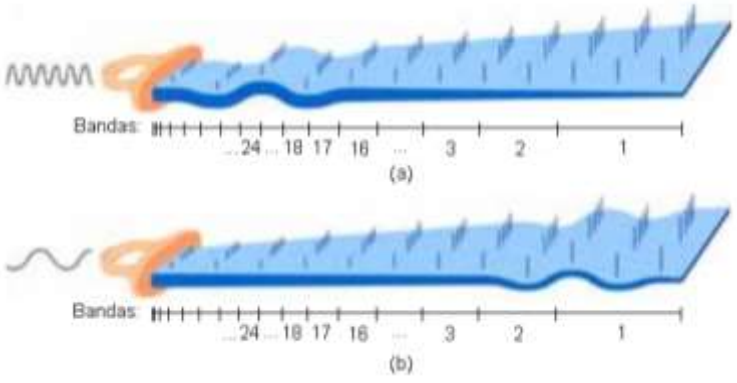


Figura 2.9 Escala de Bark y la membrana Basilar. Respuesta a frecuencias: altas (a) y bajas (b).

Esta escala tiene relación con la Escala Mel, pero no es tan conocida. Para convertir una frecuencia f a la escala Bark se utiliza la siguiente expresión:

[14] La psicoacústica estudia la percepción subjetiva de las cualidades (características) del sonido: intensidad, tono y timbre. Estas cualidades o características del sonido están, a su vez, determinadas por los propios parámetros del sonido, principalmente, frecuencia y amplitud.

$$Bark = 13\arctan\left(\frac{0.76\,f}{1000}\right) + 3.5\arctan\left(\left(\frac{f}{7500}\right)^2\right) \tag{2.16}$$

Las bandas críticas (ver tabla 2.2) son rangos de frecuencia dentro de los cuales un tono bloquea a otro tono; que ocurre cuando una onda toca la membrana basilar y la perturba dentro de una pequeña área más allá del punto de primer contacto, excitando a los nervios de toda el área vecina, por esta razón, las frecuencias cercanas a la frecuencia original no tienen mucho efecto sobre la sensación de la fuerza del sonido, aún sí doblan el volumen del sonido.

BANDA CRITICA	FRECUENCIA DE CORTE± 3 dB (Hz)	ANCHO DE BANDA (Hz)	FRECUENCIA CENTRAL(Hz)
0	0	100	50
1	100	100	150
2	200	100	250
3	300	100	350
4	400	110	450
5	510	120	570
6	630	140	700
7	770	150	840
8	920	160	1000
9	1080	190	1170
10	1270	210	1370
11	1480	240	1600
12	1720	280	1850
13	2000	320	2150
14	2320	380	2500
15	2700	450	2900
16	3150	550	3400
17	3700	700	4000
18	4400	900	4800
19	5300	1100	5800
20	6400	1300	7000
21	7700	1800	8500
22	9500	2500	10500
23	12000	3500	10500
24	15500		

Tabla 2.2 Valores para cada banda critica en la en la Escala de Bark .

La escala de Bark corresponde con la sensación auditiva asociada a la altura tonal la cual hace referencia al aspecto de la sensación auditiva por el cual podemos ordenar en una escala los sonidos que van de los bajos a los altos.

2.4.2 Coeficientes adaptados

Para realizar el análisis de las señales de voz mediante las *wavelets gaussianas complejas* aplicando el modelo auditivo humano, es necesario adaptar las bandas críticas de la escala de Bark [8] y su comportamiento logarítmico a los coeficientes de escala y traslación propicios para manipular la *wavelet* madre y generar la nueva familia con base en la estructura de análisis de frecuencias de la membrana Basilar. A continuación, se presentan los coeficientes adaptados, Tabla, 2.3

BANDAS	PARÁMETRO DE ESCALA a	FRECUENCIA DE MODULACIÓN w_0 (radianes/segundo)
1	0.002700	0.8482
2	0.002700	2.5447
3	0.002700	4.2412
4	0.002700	5.9376
5	0.002700	7.6341
6	0.002900	10.2500
7	0.002330	10.2500
8	0.001942	10.2500
9	0.001631	10.2500
10	0.001394	10.2500
11	0.001191	10.2500
12	0.001019	10.2500
13	0.000882	10.2500
14	0.000759	10.2500
15	0.000653	10.2500
16	0.000563	10.2500
17	0.000479	10.2500

Tabla 2.3 Valores de los coeficientes adaptados

Dado que la característica de frecuencia fundamental de las señales de voz no se encuentra dentro de frecuencias muy altas, solo se estudian las primeras nueve bandas; sin embargo en la Tabla 2.3 se muestran coeficientes adaptados para las primeras 17 bandas.

2.5 CREACIÓN DE LA FAMILIA *WAVELET* CGAWAVF

Seleccionada la función *gaussiana compleja* como *wavelet* madre, se procede con la creación de la nueva familia *wavelet*. Adaptando esta *wavelet* al modelo de análisis auditivo, se obtiene la nueva familia que se denominó CGAWAVF (“Complex Gaussian

Adapted *Wavelet* Family"). Para generar esta familia se necesitan los coeficientes adaptados de la escala de Bark, correspondientes al factor de escala "a" y frecuencia de modulación "w_0", que se presentaron en la Tabla 2.3

En esta forma, la función en tiempo continuo que se utiliza para realizar el análisis de las señales de voz en este proyecto queda definida de la siguiente manera:

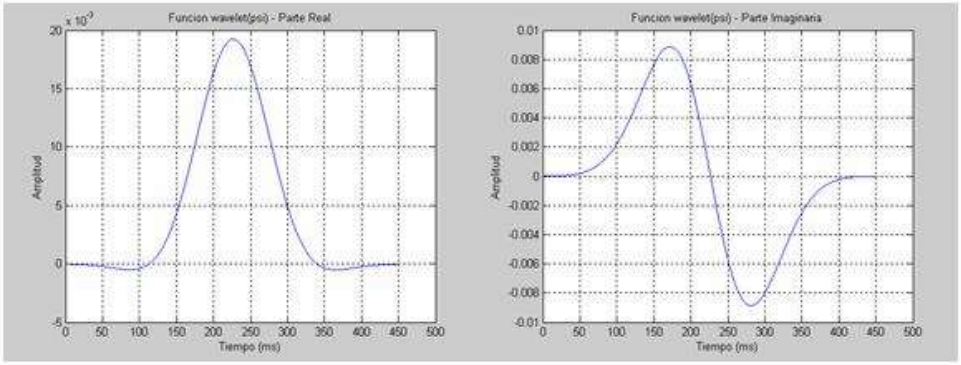
$$\psi(t) = \frac{1}{\sqrt{\pi a \sqrt{a}}} e^{-jw_0 \frac{t}{a}} e^{-\frac{1}{2}\left(\frac{t}{a}\right)^2} \quad . \qquad (2.17)$$

Mediante la discretización, la *wavelet* madre se establece como:

$$\psi_{a,w_0}(n) = \frac{1}{\sqrt{\pi a \sqrt{a}}} e^{-jw_0 \frac{n}{a}} e^{-\frac{1}{2}\left(\frac{n}{a}\right)^2} \quad , \qquad (2.18)$$

con $a \in \{$ *0.002700, 0.002900, 0.002330, 0.001942, 0.001631, 0.001394, 0.001191, 0.001019, 0.000882, 0.000759, 0.000653, 0.000563, 0.000479* $\}$; $w_0 \in \{$ *0.8482, 2.5447, 4.2412, 5.9376, 7.6341, 10.2500* $\}$ y $n \in \mathbf{Z}$.

De esta manera se creó una familia de 17 *wavelets gaussianas complejas adaptadas* a partir de la *wavelet* madre *gaussiana compleja*. A continuación se presenta la gráfica de la primera *wavelet* de esta familia, las otras *wavelet* restantes están en el Anexo B.2., que son las empleadas para el análisis de las señales de voz.



(a) (b)

Figura 2.10 *Wavelet* de la Banda 1, parámetros a = 2700X10^{-6} y w_0 =0.8482.
(a) Parte Real (b) Parte Imaginaria.

Las *wavelet*s que hacen parte de esta familia CGAWAVF en el dominio de la frecuencia están ubicadas de la siguiente forma:

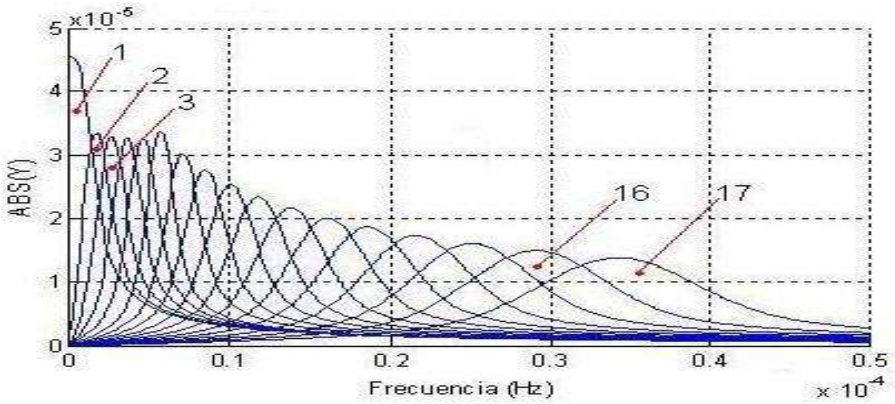


Figura 2.11 Espectro frecuencial (parte real) de las 17 *wavelet***s** de la nueva familia**.**

A continuación se presenta una gráfica donde se representa el módulo de la transformada de Fourier de las 17 *wavelet*s de la familia *CGAWAVF*.

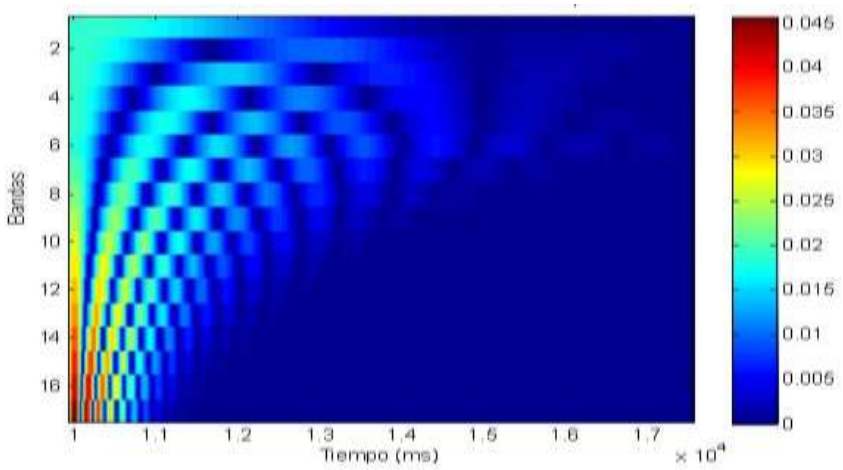


Figura 2.12 Modulo de la Transformada de Fourier de las 17 *wavelet*s de la familia CGAWAVF*.*

2.6 ESTRUCTURAS PARA EL PROCESAMIENTO DE LAS SEÑALES DE VOZ

2.6.1 *Wavelets* diádicas

Las *wavelets* diádicas son la forma sencilla de obtener la transformada *wavelet* utilizando un banco de filtros en octavas, los cuales son filtros pasa banda que dividen el

dominio frecuencial sucesivamente en dos [8]. Este tipo de análisis diádico se muestra en la figura 2.13(a).

2.6.2 Bancos de filtros

Las *wavelet*s diádicas y los bancos de filtros[15] están estrechamente relacionados [40] debido a que se puede implementar de manera eficiente la transformada *wavelet* a través de los bancos de filtros. En la figura 2.13 (b) se muestra un ejemplo de la aplicación de la transformada *wavelet* discreta con bancos de filtros de 2 canales.

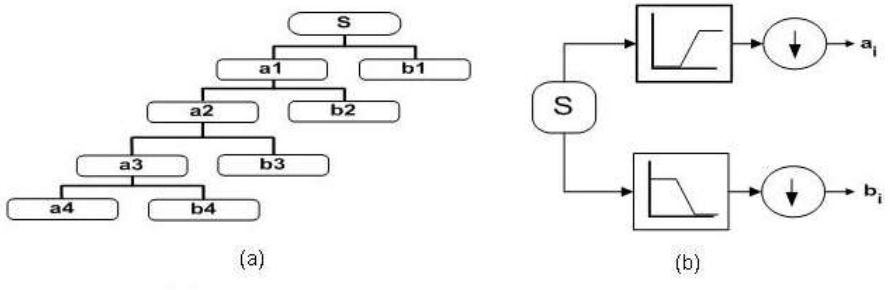


Figura 2.13 (a) Descomposición *wavelet* diádica. (b) Bancos de Filtros de dos canales

2.6.3 Estructura de bandas adaptadas.

La figura 2.13 (a) muestra que la transformada *wavelet* diádica sólo dispone de una escala para cubrir la mitad superior del recubrimiento frecuencial. Esto es, cada vez que se hace un análisis *wavelet* se divide en dos el espectro de la señal y así sucesivamente sigue ocurriendo hasta el nivel que se desee, lo que puede resultar evidentemente escaso en numerosos problemas y planteamientos [29].

En la Figura 2.14 se muestra la *estructura de bandas adaptadas* aplicada en este proyecto la cual arroja menos coeficientes y requiere menor tiempo de procesamiento que las *wavelet*s diádicas, además ésta cubre estrictamente el espectro de frecuencias que se quiere analizar. Dicha estructura aplica la transformada *wavelet* directamente en la región de las primeras bandas donde se pueden encontrar las distintas variaciones de

[15] Los bancos de filtros consisten en filtros de análisis (un filtro pasa bajas y un filtro pasa altas), un “submuestreador” y un “supmuestreador”[17].

la frecuencia fundamental de las señales de voz, de esta manera pueden tomarse hasta cinco niveles de escala, el primero con cinco bandas de análisis.

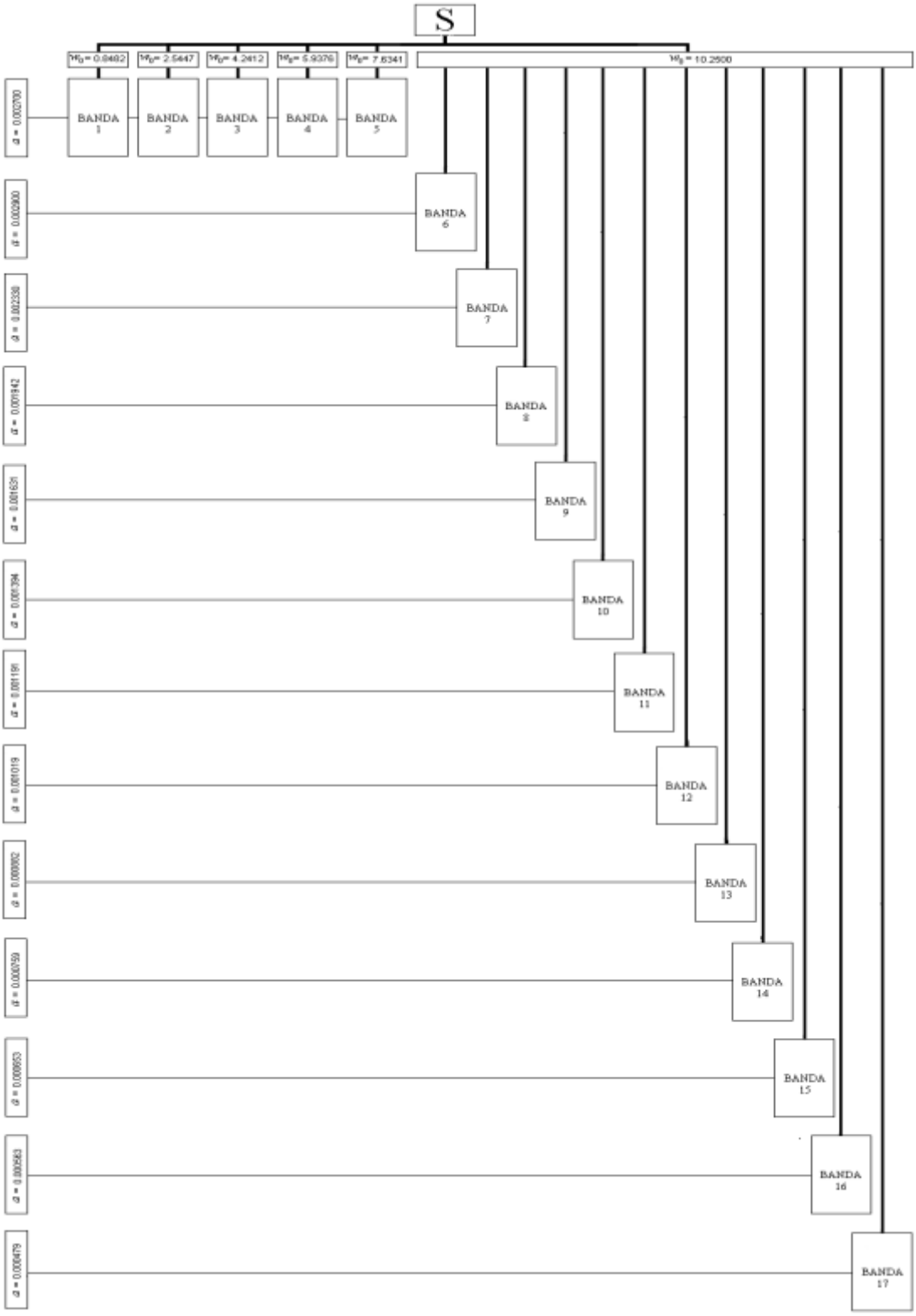


Figura 2.14 Esquema de análisis mediante la familia *wavelet cgawavf.*

Las bandas 10 hasta la 17 están diseñadas para el análisis de las frecuencias altas de las señales de voz, sin embargo no hacen parte del estudio porque las frecuencias fundamentales no se consiguen en esta región de bandas.

La señal de voz entra simultáneamente a las bandas de análisis, cada una de ellas con una *wavelet* de la familia CGAWAVF propia que tiene sus parámetros de escala a y frecuencia de modulación w_0 característicos. Este tratamiento de las señales de voz utiliza la *wavelet* madre seleccionada anteriormente (ecuación 2.18), definiendo la transformada de la siguiente manera:

$$C_{a,w_0}(t) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{\pi a \sqrt{a}}} e^{-jw_0 \frac{t}{a}} e^{-\frac{1}{2}\left(\frac{t}{a}\right)^2} dt \tag{2.19}$$

En este sentido, la ecuación 2.20 describe el proceso de cálculo de la TWD de una señal discretizada $f(n)$.

$$C_{a,w_0}(n) = f(n) * \psi_{a,w_0}(n) = \sum_{m=-\infty}^{\infty} f(n) \frac{1}{\sqrt{\pi a \sqrt{a}}} e^{-jw_0 \frac{n-m}{a}} e^{-\frac{1}{2}\left(\frac{n-m}{a}\right)^2} \tag{2.20}$$

Mediante esta transformada, cada una las bandas arrojan sus propios coeficientes de la señal procesada, que guardan características de la señal que pueden ser comunes en todas las bandas, los coeficientes son tomados posteriormente por un algoritmo para su procesamiento el cual se presentará, junto con la implementación del sistema, en el siguiente capítulo.

3. ALGORITMO DE ESTIMACIÓN DE LA FRECUENCIA FUNDAMENTAL DE SEÑALES DE VOZ CON LA FAMILIA WAVELET CGAWAVF

En este capitulo se presenta el algoritmo diseñado para que el sistema CGAWAVF realice la estimación de la frecuencia fundamental de las señales de voz de la Base de Datos VocUDC (voces de la Universidad del Cauca). Los módulos que hacen parte del sistema son: adquisición, pre-procesamiento, análisis wavelet y estimación.

3.1 FASE DE ANÁLISIS

Para la implementación del algoritmo se toma como punto de partida la información de los capítulos anteriores, de este modo el marco teórico presentado en el capítulo 1 así como el referente teórico y matemático sobre las funciones *wavelets* aplicadas a la estimación de la frecuencia fundamental de señales de voz abordado en el capítulo 2, hacen parte del estudio y análisis requerido para el desarrollo de este capitulo.

Las consideraciones identificadas como paso previo al diseño e implementación del algoritmo, son:

- Desarrollar un sistema para la estimación de la frecuencia fundamental de señales de voz mediante la aplicación de la *transformada wavelet*. El algoritmo como lo referencia Jaramillo y García [29] debe constar de unos componentes básicos como lo son: *Adquisición, Pre-procesamiento, Procesamiento Wavelet, Análisis y Estimación.*

- Establecer un entorno de experimentación en el campo de la estimación de frecuencias fundamentales de señales de voz con *wavelets* con una base de datos propia, que contenga voces del entorno regional.

- Presentar los resultados de la aplicación de la *transformada wavelet* en la estimación de la frecuencia fundamental (FF) de señales de voz en una interfaz gráfica, no sólo con el propósito de evaluar objetiva y subjetivamente los resultados, sino además visualizar posibles mejoras al sistema.

3.2 FASE DE DISEÑO

Matlab es una herramienta software empleada para el desarrollo del algoritmo. Cuenta con un toolbox [32] especializado para la aplicación de la *teoría wavelet* y que por lo tanto se utilizan algunas funciones propias de éste para desarrollar el algoritmo. De igual forma es una potente herramienta que permite desarrollar funciones propias.

3.2.1 Toolbox *wavelet* y entorno de desarrollo de interfaces gráficas de usuario

Como se menciona en [33], el toolbox *wavelet* se constituye en una gran alternativa para trabajar con *wavelets*, pues junto a la potencialidad propia de Matlab, permite realizar complejas y poderosas aplicaciones. Matlab sigue siendo entonces una de las mejores herramientas disponibles para brindar la capacidad computacional, generar datos y desplegarlos en una variedad de representaciones gráficas. Sin embargo, aunque Matlab es excelente en cálculo computacional y permite diseñar interfaces gráficas, no está diseñada para esto. Matlab tiene una interfaz de diseño grafico (GUIDE) muy pesada y algo compleja, lo que dificulta un poco su uso en comparación con otras herramientas de programación, sin embargo debido necesidades de tiempo esta herramienta resultó la más inmediata.

3.2.2 Diagrama en bloques del sistema

El sistema CGAWAVF, diseñado para estimar frecuencias fundamentales (FF) de señales de voz, esta desarrollado en módulos como se puede ver en la figura 3.1. Consta de cinco partes: *adquisición, pre-procesamiento, procesamiento wavelet, análisis y estimación.*

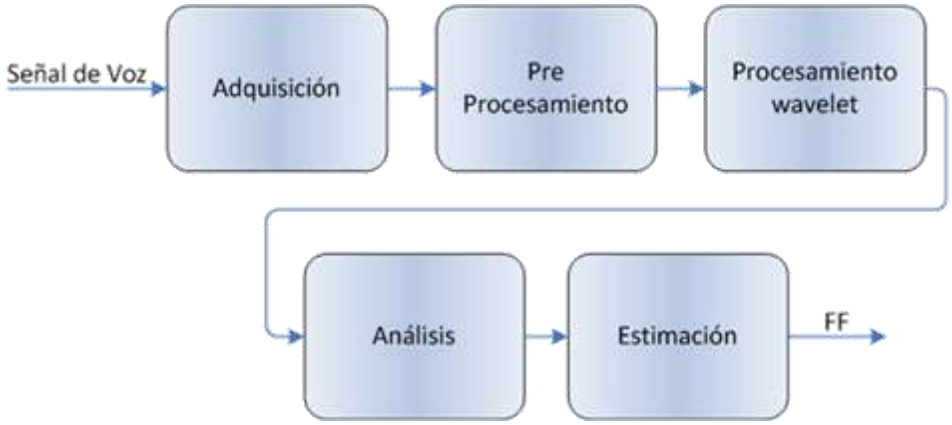


Figura 3.1 Diagrama de bloques del Sistema CGAWAVF

A continuación se presentan las consideraciones en cada uno de los módulos para el diseño del sistema.

3.2.2.1 Adquisición

En la figura 3.2 se presenta el diagrama de flujo del proceso seguido en el módulo de *Adquisición.*

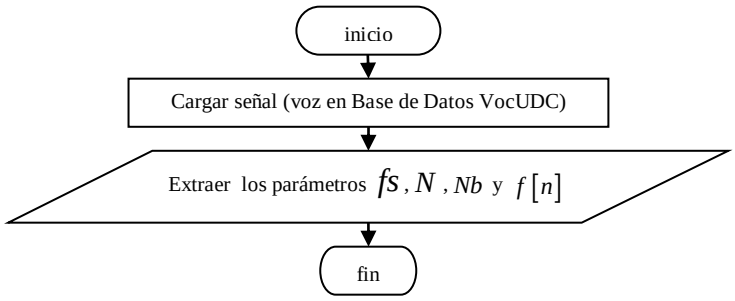


Figura 3.2 Diagrama de flujo del módulo de *Adquisición*.

El primer proceso consiste en tomar las voces de la base de datos VocUDC, que contiene los registros de voces de estudiantes de la Universidad del Cauca de diferentes procedencias (Nariño, Cauca y Valle), los cuales fueron grabados en condiciones mínimas de ruido (Set de grabación de la Emisora de la Universidad del Cauca). Cada persona realizó dos registros y cada uno de estos consta de cinco

archivos en formato WAV[16] que llevan el sonido de las vocales del alfabeto castellano (a,e,i,o,u); esto para un total de diez sonidos vocálicos por persona. Aunque la base de datos contiene señales de voz limpias de ruido, se puede operar con señales de voz ruidosas y con esto evaluar el funcionamiento del sistema con voces en otras condiciones.

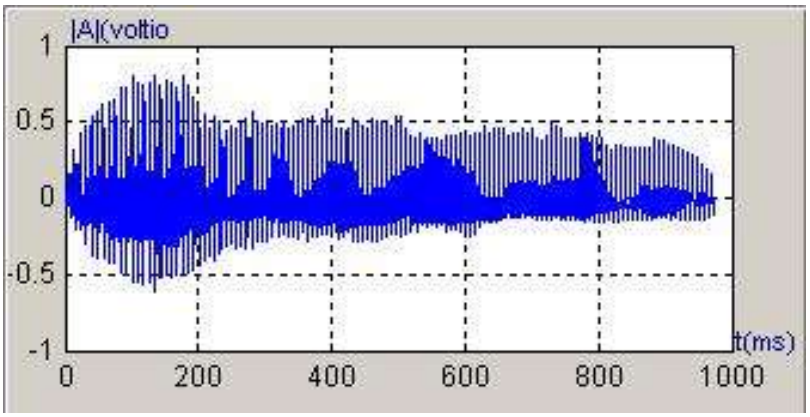


Figura 3.3 Grafica de un registro de voz (vocal a) en la base de datos VOCUDC

La segunda parte del proceso consiste en extraer los parámetros que caracterizan a cada señal de voz, los cuales están contenidos en los archivos como el nombre, la frecuencia de muestreo (fs), el número de muestras (N), el tipo de codificación de la señal (Nb) y la señal de voz misma $f[n]$. La frecuencia de muestreo y la señal son las más importantes al momento de ejecutar las etapas de pre-procesamiento, procesamiento wavelet y análisis. Estos parámetros quedan registrados en el archivo *vocal.wav* al momento en que se realizaron las grabaciones [34].

En la figura 3.3 se muestra un ejemplo de una señal de voz de la base de datos VocUDC y cuyo nombre de archivo es *a.wav*. La señal es representada por la siguiente función:

$$f[n] = \sum_{n=1}^{N=7790} x[n] \tag{3.1}$$

[16] **WAV** (o **WAVE**), apócope de *WAVEform audio format*, es un formato de audio digital normalmente sin compresión de datos desarrollado y propiedad de Microsoft y de IBM que se utiliza para almacenar sonidos en el PC, admite archivos mono y estéreo a diversas resoluciones y velocidades de muestreo, su extensión es .wav.

$$f[n] = [0.0450 \quad 0.0528 \quad 0.0495 \ldots \quad -0.0756 \quad -0.0704 \quad -0.0710], \qquad (3.2)$$

donde la frecuencia de muestreo (fs) es 8000 muestras/segundo, codificación (Nb) a 16 bits y el número de muestras (N) es 7790.

En la figura 3.4 se muestra el espectro de frecuencia de la señal que permite evaluar más adelante el compromiso frecuencial del procesamiento de la señal con el algoritmo.

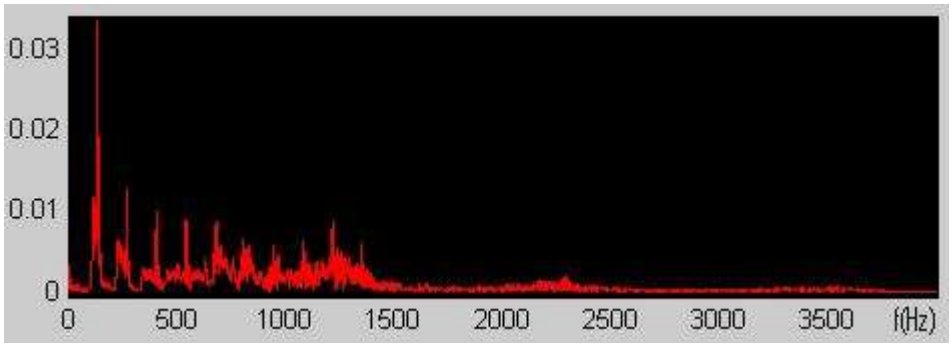


Figura 3.4 Espectro de frecuencia de la señal (vocal a).

3.2.2.2 Pre-procesamiento

El diagrama de flujo del proceso seguido en el módulo de *Pre-procesamiento* se muestra en la figura 3.5.

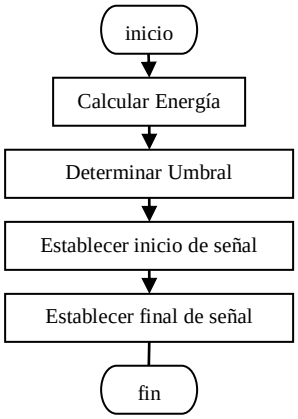


Figura 3.5 Diagrama de flujo del módulo de *Pre-procesamiento*.

El sistema está diseñado para trabajar con segmentos sonoros y no con consonantes, palabras o frases; por tanto la base de datos se hizo con vocales para garantizar la sonoridad de las señales a procesar. Sin embargo como las señales de voz fueron registradas en forma natural, no todo el tiempo de su duración mantienen la misma

amplitud; al comienzo y al final permanecen unos pequeños segmentos de señal (colas) que no son parte de la señal pero que si pueden afectar la estimación de la FF, ver figura 3.6. En resumen, esta fase consiste en establecer el inicio y final de la señal, al suprimir las colas y evitar alteraciones en la precisión de la estimación.

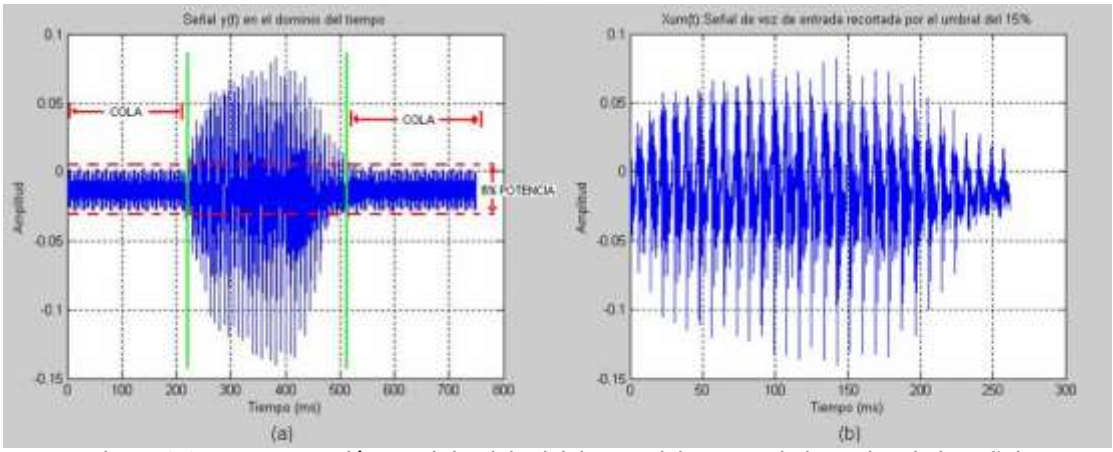


Figura 3.6 Representación en el dominio del tiempo del recorte de las colas de la señal.

Primero se calcula la energía de la señal de entrada como se muestra en la ecuación (3.3) [35]:

$$E_f = \sum \left| x[n] \right|^2 \tag{3.3}$$

Seguidamente se determina el umbral, que permite establecer los valores de la señal que son descartados, calculado a partir del 1.5% de la energía total de la señal [29]. Este porcentaje se utiliza como umbral para detectar aproximadamente dónde inicia y dónde termina la parte sonora real de la señal de voz.

$$Um = 1.5\% E_f \tag{3.4}$$

Una vez que se obtiene el umbral Um se determina el inicio de la señal a procesar, proceso que consiste en ir acumulando los valores de señal, e ir registrando la posición de esos valores hasta obtener la posición (t_i) donde se encuentra el valor de señal que completa un acumulado (A_i) igual al umbral Um. Este valor de posición (t_i) será el nuevo inicio de la señal.

Para determinar el valor de la posición donde se termina la señal, seguimos los mismos pasos del anterior proceso pero iniciando en la posición donde esta el último valor de la

señal, situando de esta manera una posición (t_f) donde el acumulado (A_f) es igual a Um:

$$f_r[n] = \sum_{t_i}^{t_f} x[n] \tag{3.5}$$

3.2.2.3 Procesamiento wavelet

En esta etapa se toma la señal pre-procesada f_r, se somete al procesamiento wavelet y se entrega una matriz ME_y con la magnitud de las señales procesadas por cada una de las bandas y un vector $Vfsc$ con sus respectivas frecuencias de muestreo. El diagrama de flujo del proceso seguido en este módulo de *Procesamiento Wavelet* se muestra en la figura 3.7.

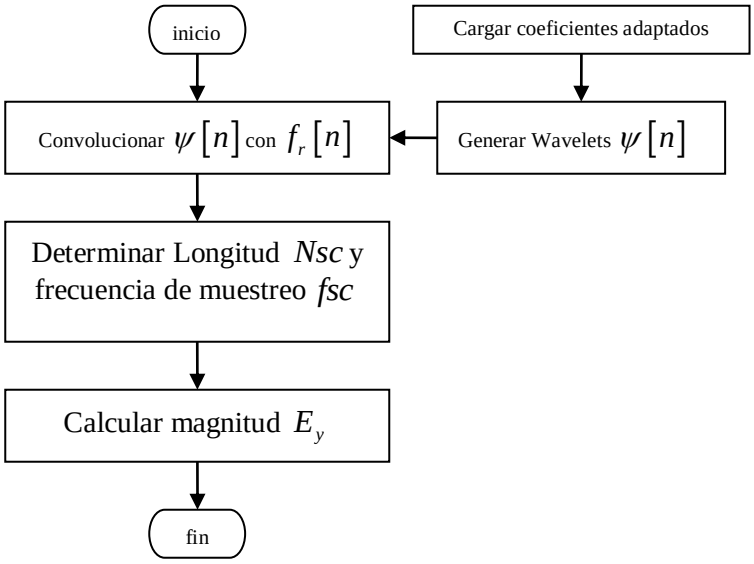


Figura 3.7 Diagrama de flujo del módulo de *Procesamiento Wavelet*.

Antes de iniciar el procesamiento, se deben cargar los coeficientes adaptados, es decir, las escalas a y las frecuencias de modulación w_0. Leonard J. García [8] muestra el proceso para la estimación de estos valores que permiten modelar las wavelets (ψ_{a,w_0}) a partir de los valores de ancho de banda y frecuencia central de cada una de las bandas criticas de la escala de Bark del modelo auditivo, el cual fue explicado en el capitulo 2 y que se ha tomado como referencia.

Luego de cargar los coeficientes adaptados, se procede a generar las *wavelets* respectivas. Se seleccionó la función Gaussiana Compleja como *wavelet* madre y se presentó el procedimiento para generar con ésta, a partir de los coeficientes adaptados, la familia CGAWAVF; un resumen de sus características de acuerdo a cada una de las bandas se muestra en la tabla 3.1.

BANDAS	PARAMETRO DE ESCALA a	FRECUENCIA DE MODULACION w_0 (rad/seg)	ANCHO DE BANDA (Hz)	FRECUENCIA DE CORTE ± 3 dB (Hz)	FRECUENCIA CENTRAL (Hz)
1	0.002700	0.8482	100	0	50
2	0.002700	2.5447	100	100	150
3	0.002700	4.2412	100	200	250
4	0.002700	5.9376	100	300	350
5	0.002700	7.6341	110	400	450
6	0.002900	10.2500	120	510	570
7	0.002330	10.2500	140	630	700
8	0.001942	10.2500	150	770	840
9	0.001631	10.2500	160	920	1000
10	0.001394	10.2500	190	1080	1170
11	0.001191	10.2500	210	1270	1370
12	0.001019	10.2500	240	1480	1600
13	0.000882	10.2500	280	1720	1850
14	0.000759	10.2500	320	2000	2150
15	0.000653	10.2500	380	2320	2500
16	0.000563	10.2500	450	2700	2900
17	0.000479	10.2500	550	3150	3400

Tabla 3.1 Características de las *wavelets* asociadas a cada una de las bandas

Por ejemplo la *wavelet* de la banda 9 que se muestra en la figura 3.8 donde aparecen su parte real (a) e imaginaria (b), se expresa según las ecuaciones (3.6) y (3.7). Teniendo en cuenta la Tabla 3.1 esta *wavelet* se caracteriza por los siguientes parámetros: escala 0.001631; frecuencia de modulación 10.2500 rad/seg; ancho de banda 160 Hz; frecuencia de corte ($\pm$ 3 dB) 920 Hz y frecuencia central 1000 Hz.

$$\psi_{a,w_0}[n] = \frac{1}{\sqrt{\pi a \sqrt{a}}} e^{-jw_0 \frac{n}{a}} e^{-\frac{1}{2}\left(\frac{n}{a}\right)^2} \tag{3.6}$$

Con $a = 0.001631;\ w_0 = 10.2500$ y $n \in \mathbf{Z}$.

$$\psi_{a,w_0}[n] = [0.0000 + 0.0001i\ldots\ -15.8076 + 13.0514i\ldots\ 0.0000 - 0.0001i] \tag{3.7}$$

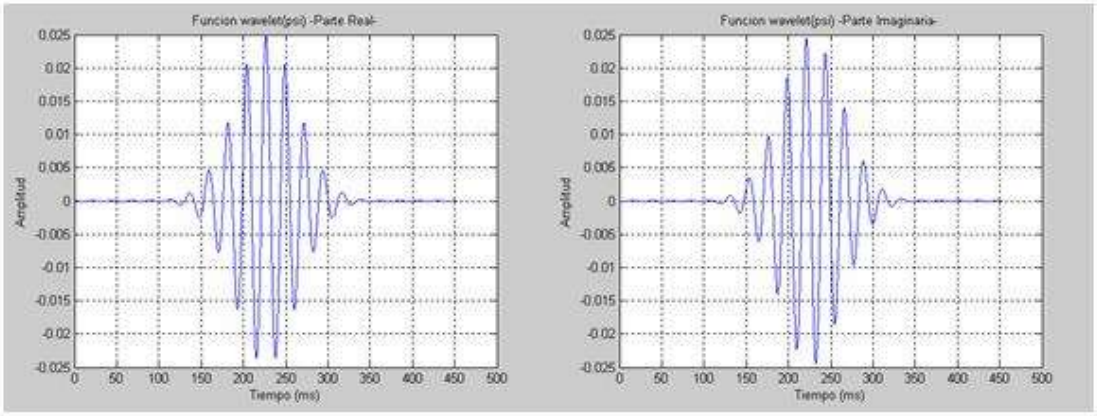


Figura 3.8 *Wavelet* de la Banda 9 (a) Parte Real (b) Parte Imaginaria.

También hay otro parámetro que caracteriza las *wavelets* y es el número de muestras N_ψ (ver Tabla 3.2). Este parámetro es muy importante porque permite determinar la longitud de la señal resultante del proceso y además determinar la frecuencia de muestreo de la misma. Una vez teniendo disponibles las *wavelets* (ψ_{a,w_0}) de cada una de las bandas se realiza el proceso de convolución de la señal pre-procesada f_r con cada una de éstas funciones.

ψ	1	2	3	4	5	6	7	8	9
N_ψ	335	335	335	335	335	281	243	215	215
ψ	10	11	12	13	14	15	16	17	
N_ψ	173	159	145	117	105	89	73	61	

Tabla 3.2 Número de muestras para cada una de las wavelets.

Por ejemplo para la banda número 9, al convolucionar la señal de voz con la *wavelet* correspondiente a esa banda, la señal procesada o resultante $y[n]$ es representada por la ecuación (3.8)[35]:

$$y[n]_9 = f_r[n] * \psi_{a,w_0}[n]_9 \tag{3.8}$$

$$y[n]_9 = [(0.0666e\text{-}4) - (0.1002e\text{-}4)i \ldots (4.1455) - (34.9505)i \ \ldots \ (0.0733e\text{-}4) + (0.1104e\text{-}4)i\ ] \tag{3.9}$$

El proceso es realizado en forma independiente, es decir, se convoluciona f_r primero con la parte real y luego con la parte imaginaria de la wavelet, con la ecuación (3.10) se realiza este proceso de la parte real y con la (3.12) el de la parte imaginaria.

$$y_{\mathrm{R}}[n]_9 = f_r[n] * \mathrm{Re}\left\{\psi_{a,w_0}[n]_9\right\} \tag{3.10}$$

$y_R[n]_9$ = [0.0666e-4 0.1590 e-4 …4.1455 19.1111… 0.2085 e-4 0.0733 e-4] (3.11)

En la figura 3.9 se muestra el resultado de la ecuación (3.10) donde se procesa la señal de entrada con la parte real de la wavelet de la banda 9.

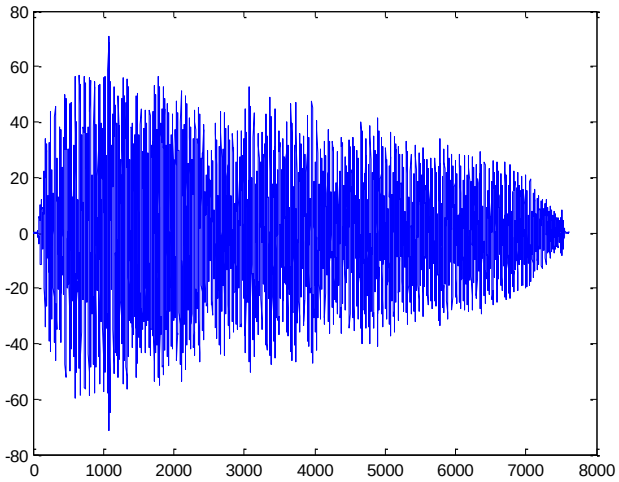


Figura 3.9 Procesamiento de la señal de entrada con la parte real de la *wavelet* de la banda 9.

De igual forma en la figura 3.10 se muestra el resultado de la ecuación (3.12) donde se procesa la señal de entrada con la parte imaginaria de la *wavelet* de la banda 9.

$$y_I[n]_9 = f_r[n] * \mathrm{Im}\left\{\psi_{a,w_0}[n]_9\right\} \tag{3.12}$$

$y_I[n]_9$ = [0.1002 e-4 0.1129 e-4… 34.9505 28.6377… -0.1746 e-4 -0.1104 e-4] (3.13)

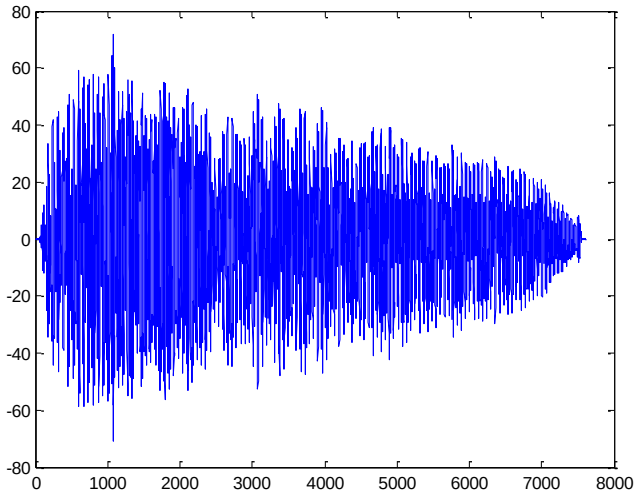


Figura 3.10 Procesamiento de la señal de entrada con la parte imaginaria de la wavelet de la banda 9.

Seguidamente se determina el número de muestras Nsc y la frecuencia de muestreo fsc. El número de muestras de esta señal resultante $y(n)$ se calcula según la ecuación (3.14)[35]. En el caso del ejemplo anterior, la señal resultante en las ecuaciones (3.10) y (3.12) tienen la misma longitud, 7614 muestras en total.

$$Nsc = (N_{\psi} + N) - 1, \tag{3.14}$$

donde N es la longitud de la señal de voz y N_{ψ} es la longitud de la wavelet. La frecuencia de muestreo de la señal resultante fsc se obtiene como se muestra en la ecuación (3.15), necesaria para posteriormente obtener la frecuencia fundamental.

$$fsc = \left(\frac{N_{\psi} + 1}{N} \right) fs\,. \tag{3.15}$$

Finalmente se calcula la magnitud de la señal resultante E_y mediante la ecuación (3.16) [35]. La grafica de la Magnitud de la señal procesada por la wavelet de la banda nueve puede verse en la figura 3.11.

$$E_y[n] = \sqrt{(y_R[n])^2 + (y_I[n])^2} \tag{3.16}$$

$E_y[n]$ = [0.1203e-4 0.1950e-4… 35.1955 34.4289… 0.2719 e-4 0.1326 e-4] (3.17)

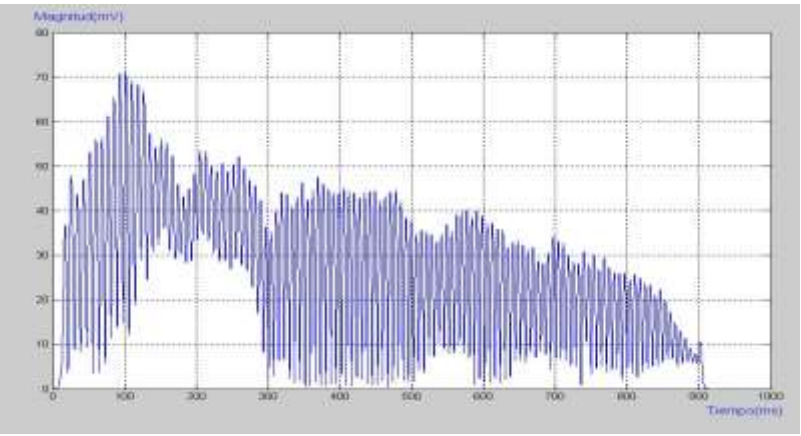


Figura 3.11 Grafica de la magnitud de la señal procesada por la *wavelet* de la banda 9.

El mismo proceso se hace para cada una de las bandas con las que se procesó la señal de voz. Si denominamos E_{yi} a la señal obtenida de procesar la señal de voz en la banda i y a fsc_i la frecuencia de muestreo de esa señal resultante; con $i = 1,2,3...17$, se pueden organizar los vectores E_y en una matriz ME_y y los valores fsc en un vector $Vfsc$. De esta forma las dimensiones de la matriz ME_y son [$N_{\max}$ x17] y $Vfsc$ [1 x 17], donde $N_{\max}$ corresponde a la longitud máxima de las señales resultantes del procesamiento wavelet. Puesto que todos los vectores E_y no van a quedar de la misma longitud, debido a las diferentes longitudes de las *wavelets*, a los vectores E_y de menor longitud se les hace un relleno con ceros. A continuación se muestra una representación de la matriz ME_y.

$$ME_y = \begin{bmatrix} E_{y1} \\ \dots \\ E_{y5} \\ E_{y6} \\ E_{y7} \\ \dots \\ E_{y17} \end{bmatrix} = \begin{bmatrix} e_{y11} & \dots & e_{y1N_{\max}-274} & \dots & e_{y1N_{\max}-92} & \dots & e_{y1N_{\max}-54} & \dots & e_{y1N_{\max}-1} & e_{y1N_{\max}} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ e_{y51} & \dots & e_{y5N_{\max}-274} & \dots & e_{y5N_{\max}-92} & \dots & e_{y5N_{\max}-54} & \dots & e_{y5N_{\max}-1} & e_{y5N_{\max}} \\ e_{y61} & \dots & e_{y6N_{\max}-274} & \dots & e_{y6N_{\max}-92} & \dots & e_{y6N_{\max}-54} & 0 & \dots & 0 \\ e_{y71} & \dots & e_{y7N_{\max}-274} & \dots & e_{y7N_{\max}-92} & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ e_{y171} & \dots & e_{y17N_{\max}-274} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (3.18)$$

$$Vfsc = \begin{bmatrix} fsc_1 & fsc_2 & fsc_3 & . & fsc_i & . & fsc_{17} \end{bmatrix} \quad (3.19)$$

Donde cada fsc_i corresponde a un E_{yi}. Además, se puede observar en la Tabla 3.2 que las E_y de las primeras cinco bandas tienen igual tamaño y son siempre las de mayor longitud.

3.2.2.4 Análisis

El diagrama de flujo del proceso seguido en este módulo se muestra en la figura 3.12.

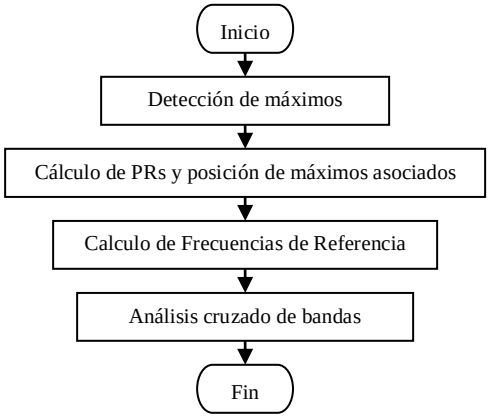


Figura 3.12 Diagrama de flujo del módulo de *Análisis*.

En esta etapa, el primer paso es la detección de los máximos ($e_{y\max}$), para cada una de las bandas, a partir de la matriz ME_y. Por ejemplo, al realizar una ampliación entre 300 y 500 milisegundos sobre la señal resultante de la banda 9, se pueden ver claramente 28 máximos marcados con puntos de color verde (ver figura 3.13). El procedimiento de detección consiste en determinar la posición de los máximos, evaluar su magnitud y seleccionar aquellos que superen el porcentaje de umbral de magnitud $U\max$. Para una banda (i) con una señal resultante ($E_y i$) y un número total de (n) máximos, se conforma un vector $Ve_{yn}i$ de la forma:

$$Ve_{yn}i = \left[e_{y1}i \quad e_{y2}i \quad \ldots \quad e_{yn-1}i \quad e_{yn}i \right] \tag{3.20}$$

Experimentalmente se determinó tener en cuenta para cada banda a todos los máximos encontrados por encima del 0.5% del promedio de las magnitudes y descartar aquellos máximos pequeños que puedan afectar el cálculo de la FF.

$$U\max = \frac{0.005}{n}\sum_{j=1}^{n} e_{yj} i \tag{3.21}$$

El número total de máximos considerados es m y se representan por el siguiente vector:

$$Ve_y i_{\max} = \left[e_{y1} i_{\max} \quad ... \quad e_{yj} i_{\max} \quad ... \quad e_{ym} i_{\max} \right] \tag{3.22}$$

con $e_{yj} i_{\max} \geq Umax$ *y* $m \in \mathbf{Z}^{+} \leq n$.

Seguidamente se calculan los periodos de referencia (PR) y las posiciones de los máximos asociados. Un PR corresponde a la separación que existe entre dos máximos consecutivos ($e_{yj} i_{\max}$ y $e_{yj+1} i_{\max}$), por tanto para una cantidad de m máximos se tienen $m-1$ periodos de referencia PRs (ver figura 3.13).

Cada PR debe estar asociado a un máximo, ésto se debe a que el análisis para la estimación de la FF esta basado en esta correspondencia. Para asociar el PR se puede tomar el máximo que ocurra ya sea de primero o de segundo, teniendo en cuenta que al seleccionar el que ocurre primero se descarta el último máximo y si se asocia el PR al segundo se descarta el primer máximo. Para este trabajo se asociaron los PRs al máximo que ocurrió primero; por tanto el ultimo máximo no va a tener PR asociado. De esta forma, de cada dos máximos consecutivos se obtiene un PR, generando dos vectores V_{POS} $(1\ X\ m-1)$ y V_{PR} $(1\ X\ m-1)$, uno con la posición de los máximos que tienen asociados un PR y el otro con dichos PRs.

$$V_{POS} = \left[\ POS_1 \quad POS_2 \quad POS_3 \quad ... \quad POS_i \quad ... \quad POS_{m-1} \right] \tag{3.23}$$

$$V_{PR} = \left[\ PR_1 \quad PR_2 \quad PR_3 \quad ... \quad PR_i \quad ... \quad PR_{m-1} \right] \tag{3.24}$$

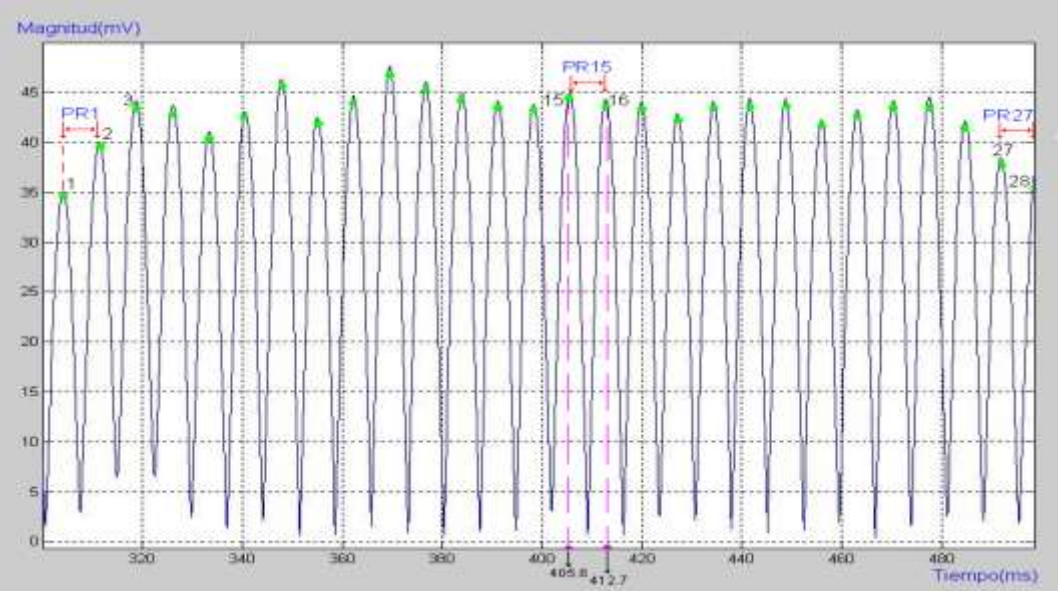


Figura 3.13 Grafica de máximos y PRs en un segmento de la señal procesada por la wavelet de la banda 9.

Al realizar el proceso para todas las bandas se obtienen M_{PR} y M_{POS}.

$$M_{PR} = \begin{bmatrix} V_{PR}1 \\ \dots \\ V_{PR}5 \\ V_{PR}6 \\ V_{PR}7 \\ \dots \\ V_{PR}17 \end{bmatrix} \quad (3.25) \qquad M_{POS} = \begin{bmatrix} V_{POS}1 \\ \dots \\ V_{POS}5 \\ V_{POS}6 \\ V_{POS}7 \\ \dots \\ V_{POS}17 \end{bmatrix} \quad (3.26)$$

El paso a seguir en este proceso es el cálculo de las Frecuencias de Referencia (FR). A partir de las matrices de posición de los máximos y su matriz de PRs asociados se calculan las FRs. Los vectores V_{FR} y la matriz M_{FR}, con los valores de FRs, se obtienen como se muestra en (3.27).

$$M_{FR} = \begin{bmatrix} V_{FR}1 \\ \dots \\ V_{FR}5 \\ V_{FR}6 \\ V_{FR}7 \\ \dots \\ V_{FR}17 \end{bmatrix} = \begin{bmatrix} fsc_1 / V_{PR}1 \\ \dots \\ fsc_5 / V_{PR}5 \\ fsc_6 / V_{PR}6 \\ fsc_7 / V_{PR}7 \\ \dots \\ fsc_{17} / V_{PR}17 \end{bmatrix} \quad (3.27)$$

Cada fila de esta matriz M_{FR} corresponde a los datos obtenidos del procesamiento wavelet de la señal por una banda específica.

Como último paso de este módulo, se procede con el *Análisis Cruzado de Bandas.* Al Tomar FR_j1, del vector $V_{FR}1$ de la matriz M_{FR}, se empieza a hacer el cruce con los demás $FR_j i$ que estén alrededor de POS_j en las demás bandas. Esto es, se establece FR_j1 como una ventana sobre los máximos que estén localizados en POS_j y se toman todos los $FR_j i$ asociados a esos máximos que caigan dentro de la ventana, ver figura 3.14.

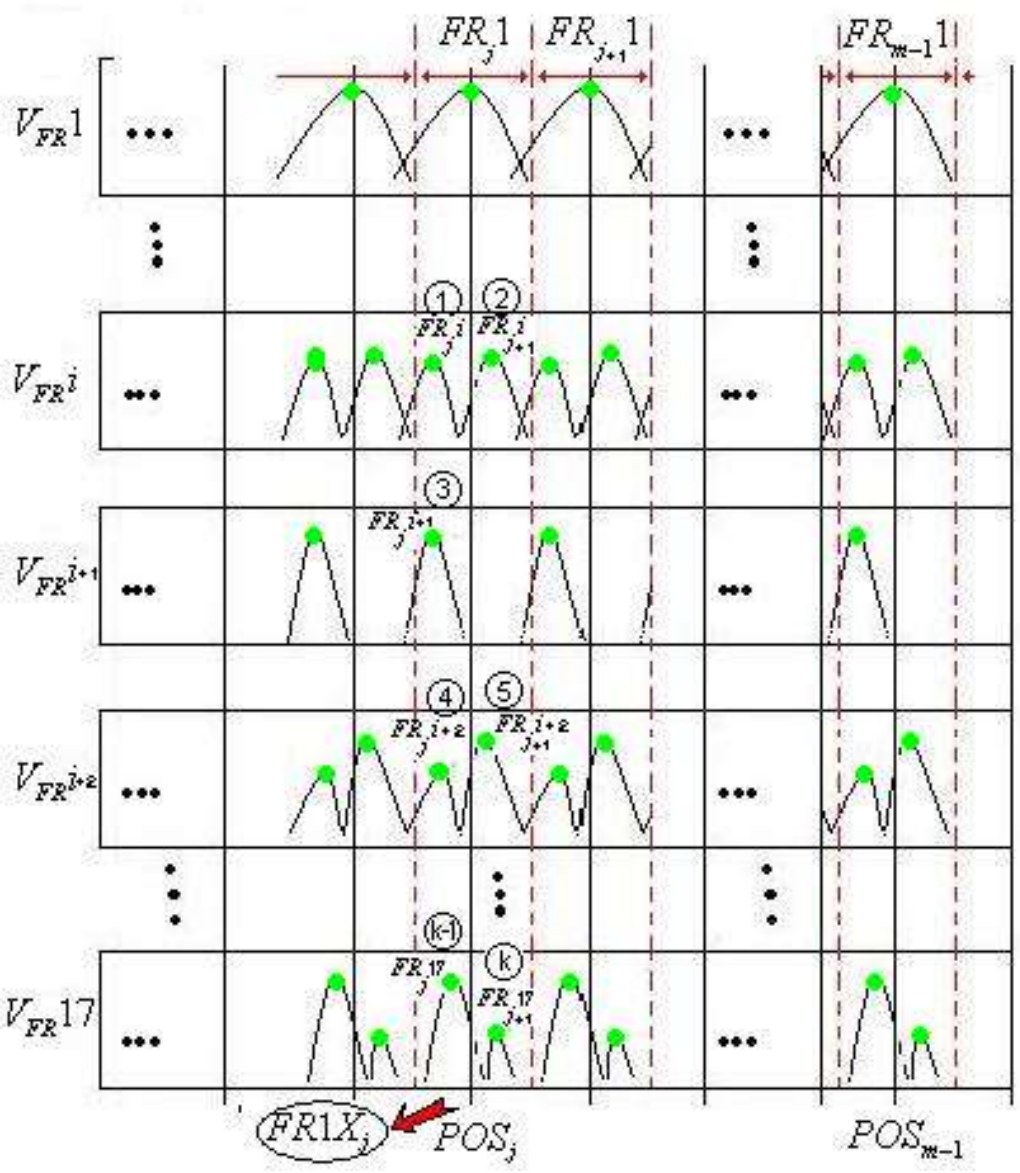


Figura 3.14 Grafica de análisis cruzado de bandas tomando como referencia a FR_j1, del vector $V_{FR}1$.

Como se puede observar en la figura 3.14 se obtienen k valores $FR_j i$ que están dentro de la ventana FR_j1 localizada sobre la posición POS_j, de ellos se determina el valor que mayor número de veces se presente ($FR1X_j$). Luego se repite el procedimiento tomando como referencia $FR_{j+1}1$ de donde da como resultado otro $FR1X_{j+1}$ y así sucesivamente con todos los FR_j1 del vector $V_{FR}1$ hasta obtener un vector $V_{FRX}1$ con los valores $FR1X$.

$$V_{FRX}1 = \left[FR1X_1 \quad ... \quad FR1X_j \quad ... \quad FR1X_{m-1} \right] \tag{3.28}$$

El anterior procedimiento se repite, pero tomando como referencia $V_{FR}2$ y así sucesivamente hasta $V_{FR}17$ de la matriz M_{FR} para, finalmente, obtener una matriz M_{FRX} así:

$$M_{FRX} = \begin{bmatrix} V_{FRX}1 \\ ... \\ V_{FRX}5 \\ V_{FRX}6 \\ V_{FRX}7 \\ ... \\ V_{FRX}17 \end{bmatrix} \tag{3.29}$$

3.2.2.5 Estimación

El ultimo módulo, cuyo diagrama de flujo se muestra en la figura 3.15, se presenta a continuación.

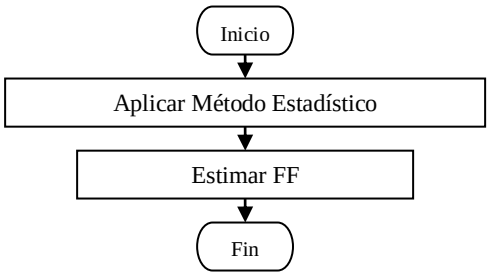


Figura 3.15 Diagrama de flujo del módulo de *Estimación*.

Como primera parte de esta fase se procede con la aplicación del método estadístico adaptado [36][37]. Este método, adaptado a los valores de frecuencias fundamentales de las voces registradas en la base de datos VocUDC (Anexo C.2), arroja las medidas de tendencia central, media, mediana y moda, de cada uno de los vectores de banda las cuales son organizadas en la matriz M_{TC} .

$$M_{TC} = \begin{bmatrix} Med1 & Mdna1 & Mda1 \\ \dots & \dots & \dots \\ Med2 & Mdna2 & Mda2 \\ Med3 & Mdna3 & Mda3 \\ Med4 & Mdna4 & Mda4 \\ \dots & \dots & \dots \\ Med17 & Mdna17 & Mda17 \end{bmatrix} \tag{3.30}$$

El paso final de este módulo y de todo el algoritmo es la estimación de la FF a partir de la matriz M_{TC}. De los valores de *moda* obtenidos (columna 3 de la matriz M_{TC}), se procede a determinar nuevamente la media, mediana y ***moda***, este último valor ($MdaX$) es el que se toma como la frecuencia fundamental estimada, dado que es el que arroja las mejores estimaciones [8].

3.3 FASE DE IMPLEMENTACIÓN

La fase de implementación del algoritmo implica el desarrollo de las funciones en Matlab a través de las cuales se da funcionalidad a cada uno de los módulos diseñados en la fase anterior. Algunas de estas funciones operan en conjunto con funciones proporcionadas por el *toolbox wavelet*, sin embargo, estas últimas no son descritas en detalle pues basta con emplear el *help* de Matlab para obtener más información sobre ellas. Considerando además que la implementación total del algoritmo abarca una gran cantidad de funciones, se describe a continuación únicamente las más importantes, sobre todo, las que implementan directamente los módulos en la fase de diseño.

La implementación del sistema de estimación de la FF está codificada, por practicidad, en archivos *.m y* organizada en funciones de procesamiento y funciones de análisis. Para la implementación del algoritmo se crearon algunas funciones, otras fueron tomadas del Toolbox de Matlab y otras fueron modificadas a partir de las disponibles en el Toolbox.

3.3.1 Funciones de procesamiento

Las funciones de procesamiento se encargan de procesar las señales de voz.

ProcesarSxB. Esta función toma la señal de voz ($vcal$ - cualquiera de las vocales-), la frecuencia a la que fue muestreada la señal (fs) y el rango de bandas para el procesamiento determinado por la banda inicial (bi) y final (bf). La sintaxis de la función es: $\left[MEyo \quad Fsc \quad MYori\right] = procesarSxB\left(vcal, fs, bi, bf\right)$. Respectivamente la función entrega una matriz que contiene las señales procesadas ($MEyo$) en el rango de bandas seleccionadas y un vector con las frecuencias de muestreo (Fsc) de cada una de estas señales; también entrega otra matriz con las señales resultantes en forma compleja ($MYori$).

Esta función se apoya en las siguientes funciones:

- Escalamod: Contiene una matriz con los valores de escala (a) y frecuencia de modulación (w_0) adaptados. La sintaxis de la función que permite cargar estos valores es: *load ('escalaamod.m'), load* es una función del Toolbox de Matlab.

- Cgamwavf: Permite generar las wavelets teniendo en cuenta a *escalaamod (eam)* y la banda (*b*) que se quiera generar. La sintaxis de la función es: *[psi a* w_0 *fsp] = cgamwavf(eam,b)*. Los parámetros de salida arrojan la wavelet en formato complejo (*psi*), el factor de escala (*a*), la frecuencia de modulación (w_0), la frecuencia de muestreo (*fsp*) de la wavelet (20000 Hz).

- Estimffwavfcgam: Esta función realiza la convolución de la señal de entrada (*vocal*) con una wavelet predeterminada (*psi*). Requiere como parámetro de entrada la frecuencia de muestreo de la voz a procesar (*fs*). La sintaxis de esta función es: *[Yor,Yoi,Xum,fsc] = estimffwavcgam(psi,vocal,fs)*. *Fsc* es la frecuencia de muestreo de la señal resultante; *Xum* es la señal original recortándole las “colas” y *Yor* e *Yoi* son la parte real y la parte imaginaria de la señal resultante, respectivamente.

3.3.2 Funciones de análisis

Este grupo de funciones toman los resultados del procesamiento y realizan el estudio de los máximos para la estimación de la frecuencia fundamental. Esta se divide en tres partes: análisis de señal, análisis estadístico y estimación.

Análisis de señal.

El Análisis de señal comprende el estudio de la señal para la detección y extracción de los máximos de las señales procesadas y sus funciones son:

- Detectamax: Esta función permite determinar los máximos de la señal procesada por cada una de las bandas. La sintaxis de la función es *[maxis Emaxis]=detectamaxE(Eyo)*. Necesita como parámetro de entrada la energía de la señal procesada (*Eyo*). Arroja la posición de los máximos (*maxis*) y sus valores de energía (*Emaxis*).

- PfrefE: Función para hallar los periodos de referencia (*pref*) asociados a cada máximo tomando como dato de entrada el vector de máximos locales (*max*) de una señal. La sintaxis es *[pref]=pfrefE(maxs)*.

- GenMpseccband: Función que a partir de la matriz de posición de los máximos en cada banda (*Mpos*), la matriz de periodos de referencia (*Mpit*) y tomando una banda predeterminada como referencia (*bda*), genera la matriz con los valores de período fundamental (*Mpsecct*) que se encuentran dentro los valores de periodo fundamental de referencia en el vector de referencia. De igual forma se obtiene un vector que contiene los valores de periodo fundamental que mayor número de veces se presenta (*vpseccmoda*) en cada una de las bandas de *Mpsecct*. La sintaxis de la función es *[Mpsecct,vpseccmoda]=genMpseccband(Mpos,Mpit,bda)*.

Análisis Estadístico.

El análisis estadístico abarca la implementación de funciones que aplican un método estadístico adaptado (ver Anexo C.2) a datos del tipo frecuencia fundamental de señal de voz y consta de las siguientes funciones:

- Frecsimple: Aplica la distribución de frecuencias simple al vector de valores de períodos fundamentales *vpseccmoda* (*vp*). Retorna la tabla con la distribución de frecuencias simple (*dot*) y el valor mínimo (*linf*) y máximo (*lsup*) que se presenta en el vector *vp*. La sintaxis de esta función es *[dot linf lsup]=frecsimple(vp)*.

- Contar: La sintaxis de esta función es *[c,n,Am,mr,R]=contar(vp)*. Aplica la distribución de frecuencias en intervalos al vector de valores de periodo fundamental *vp*. Retorna la tabla con la distribución de frecuencias (*c*) que contiene los intervalos (*Limite inferior* y *Limite Superior*), marca de clase (*X*), frecuencia en cada intervalo (*F*), frecuencia acumulada (*Fa*), frecuencia relativa (*Fr*) y frecuencia relativa acumulada (*Fra*); número total de datos de *vp* (*n*); amplitud definitiva de los intervalos (*Am*); número total definitivo de intervalos (*mr*) y el rango de los datos de *vp* (*R*).

- Tendcentral: La función es *[vtendc]=tendcentral(c,n,Am)* que calcula las medidas de tendencia central media, mediana y moda de la distribución de frecuencias en intervalos (*c*). Retorna el vector (*vtendc*) con el número de la vocal analizada, la media aritmética, la mediana y la moda de los datos de *c*.

- Meddisper: Calcula las medidas de dispersión: rango (*R*), desviación media (*Dm*), varianza (*S2*), desviación estándar (*S*) y coeficiente de variabilidad (*Cv*) de los datos de *c*. Retorna la matriz de datos (*D*) de donde se calculan de las medidas de dispersión y el vector con las medidas de dispersión (*RDmS2SCv*). La sintaxis de la función es *[D,RDmS2SCv]=meddisper(c,vtc,R)*.

Estimación.

La estimación corresponde al método que permite determinar el valor que más se presenta dentro de los valores obtenidos en el análisis estadístico de cada banda.

Por medio de las funciones *mean* (*Mtc*), *median* (*Mtc*), *mode* (*Mtc*) del toolbox de Matlab se calculan las medidas de tendencia central de los valores de frecuencia fundamental más frecuentes arrojados por cada una de las bandas cruzadas. *Mtc* es la matriz que contiene todos los *vtc* de cada una de las bandas. El valor que arroja la función *mode* es el que se considera como la frecuencia fundamental estimada [8].

3.4 CARACTERÍSTICAS

Las características de implementación del algoritmo desarrollado se presentan en los siguientes puntos.

- Es un algoritmo de estimación de frecuencias fundamentales de señales de voz, diseñado inicialmente para procesar señales no ruidosas, que aplica la técnica *wavelet*. El algoritmo se ha trabajado con las señales de voz de la base de datos VocUDC las cuales han sido registradas bajo condiciones mínimas de ruido.

- Implementa la DWT a partir de un modelo auditivo real. Las *wavelets* se han adaptado al funcionamiento real de la membrana Basilar que hace parte del sistema auditivo humano.

- La familia de *wavelet* empleada (Gaussiana Compleja) se utiliza específicamente para análisis de las señales de voz y no está adecuada para la síntesis de éstas.

- El sistema de estimación se basa en el método estadístico que ha sido adaptado para analizar valores en los rangos de frecuencia donde se encuentran las frecuencias fundamentales de la voz. La moda, es la medida de tendencia central sobre la que se fundamenta la estimación de la frecuencia fundamental.

Entre las principales características funcionales del sistema se encuentra que permite visualizar las señales de voz en el dominio del tiempo y frecuencia durante el procesamiento *wavelet*. Además puede generar un reporte de los valores de tendencia central de donde se toma la frecuencia fundamental estimada para cada procesamiento que se realice. También se pueden visualizar las *wavelets* utilizadas en cada banda de procesamiento.

3.5 MODO DE EJECUCIÓN DEL ALGORITMO

Para la ejecución del algoritmo implementado se tiene dos posibilidades. La primera de ellas es acceder a la GUI ejecutando el comando *guide* en el prompt, lo que despliega

una interfaz propia de Matlab que permite abrir una GUI existente seleccionando la opción *'Open Existing GUI'*. Mediante el browser se selecciona el archivo *algoritmo_wavelets.fig* y se pone en funcionamiento dando clic en el botón '*Run*' ubicado en la parte superior del *Layout Editor*.

La segunda posibilidad consiste en ejecutar en el prompt la función cuyo nombre es el mismo con el que se identifica al archivo .fig generado en el proceso de desarrollo mediante GUIDE, es decir, basta con ejecutar el comando *algoritmo_wavelets* para inicializar al algoritmo, siempre y cuando el *path* de trabajo de Matlab esté direccionado al directorio de la aplicación. En el anexo C.3 se presenta a manera de manual de usuario detalles del empleo del algoritmo.

3.6 LIMITACIONES

Se presenta variabilidad en la precisión de la estimación de la frecuencia fundamental si se utilizan registros de voces que estén grabadas en condiciones de ruido. Cualquier otro tipo de señal que se utilice debe tener en cuenta las condiciones mínimas de ruido y los criterios adecuados para la grabación como: equipos, cabina, pronunciación y todos los aspectos considerados necesarios en la realización de la base de datos VocUDC, como los que se van a mencionar en el capitulo 4.

4. SISTEMA CGAWAVF, LA VocUDC Y EL SPEECH FILING SYSTEM

Este capitulo contiene los resultados obtenidos de estimar la *frecuencia fundamental* FF, en registros de voz bajo el sistema CGAWAVF y la confrontación de su desempeño con el del software "*Speech Filing System*". Los registros que fueron grabados en las instalaciones de la emisora de la Universidad del Cauca corresponden a individuos hombres y mujeres entre los 19 y 28 años de edad del suroccidente colombiano específicamente de ciudades de los departamentos de Nariño, Cauca y Valle. Este material auditivo se almacenó en la base de datos **VocUDC**, importante para no perder dichos registros vocálicos, materia prima que puede servir en futuras investigaciones.

4.1 BASE DE DATOS DE VOCES DEL SUROCCIDENTE COLOMBIANO

La base de datos ***VocUDC*** (Voces del suroccidente colombiano de la Universidad del Cauca), ha sido elaborada para el proyecto de procesamiento de señales de voz con técnica *wavelet*, con el apoyo de los departamentos de Ingeniería Electrónica y Telecomunicaciones, el departamento de Fonoaudiología y la División de Comunicaciones de la Universidad del Cauca.

Esta base de datos contiene los registros grabados de las señales vocálicas emitidas por individuos del suroccidente colombiano de habla hispana. Se tiene la seguridad que estas grabaciones son un material muy importante para quienes quisieran profundizar en dicha investigación tanto desde el campo del procesamiento digital de señales bajo la técnica *wavelet* como desde la orientación médica a la fonoaudiología.

Para que la creación de esta base de datos tenga el valor que se espera, el procedimiento de grabación debe ser lo más fiel posible a la señal misma y así, se asegure un buen grado de confiabilidad en el material colectado y éste pueda ser un apoyo confiable en futuras investigaciones.

4.1.1 Anamnesis de Voz [1][2]

El proceso de grabación se inicia seleccionando hablantes del suroccidente colombiano, estudiantes de las Facultades de Ingenierías de la Universidad del Cauca, muestra que se tomó teniendo en cuenta que de la población universitaria estos individuos responderían a un grupo con un registro de voz que pueda catalogarse como “normal”, cumpliendo con características como por ejemplo: que no sean fumadores excesivos, que tengan un uso no excesivo de su voz y que no posean enfermedades asociadas al aparato fonatorio, entre otras.

La selección de los individuos se realizó teniendo en cuenta un diagnóstico previo apoyado por la especialista en el estudio de la voz Miryam Adela Barreto profesora del programa de Fonoaudiología, de la Facultad de Ciencias de la Salud de la Universidad del Cauca, quien acompañó las diversas actividades realizadas para la construcción de la base de datos y especialmente en el análisis y adecuamiento del documento tipo encuesta, propio del Departamento de Fonoaudiología, para el estudio de la voz, llamado anamnesis del área del habla.

Ésta encuesta consiste en recolectar información que permita la identificación del sujeto, antecedentes foniátricos, conductas de esfuerzos, modificación de la voz, condiciones en que ejerce su actividad, tiempo de desarrollo de las disfonías, síntomas asociados con las disfonías, problemas de salud en general, tratamientos médicos efectuados, aspecto psicológico, entre otros.

Del documento se toman apartes y se le realizan las modificaciones que responden a las necesidades del proyecto; esta encuesta permite dilucidar los parámetros físicos y fisiológicos que alteran la voz, datos útiles para determinar a groso modo si el hablante pertenece o no a una población con voz “normal”. El cumplir con cualquiera de las características evaluadas como negativa en la encuesta para una voz normal, fue el condicionante para la realización de su registro.

El aporte que realiza al proyecto la docente desde su área de trabajo permite tener un criterio apropiado en el momento de evaluar la encuesta empleada para determinar la población a muestrear, seguidamente se lleva a cabo una prueba piloto para verificar si

dicha información respondía a las necesidades del proyecto, dicha prueba realizada en la Facultad de Derecho permitió modificar y enriquecer el instrumento (encuesta) empleado para colectar la información necesaria para seleccionar la población muestra.

El resultado del anterior proceso es el documento de anamnesis con un instructivo para realizar una buena evaluación que permita descartar la presencia o no de algún hablante que pertenezca a la población–riesgo y que por las cualidades del proyecto no podría pertenecer a la base de datos ***VocUDC*** de la Universidad del Cauca.

Para la facilidad y mejor entendimiento del evaluado y el evaluador, se dividió este nuevo documento en tres partes: datos generales, hábitos y antecedentes. A continuación se dará una relación de cada uno de ellos, con sus respectivas explicaciones, lo que permitirá a futuras investigaciones sobre estas señales de voz adelantar en el camino del diagnostico preliminar de los hablantes, e identificar a aquellos que son idóneos para realizar las grabaciones y pruebas de sus voces.

4.1.1.1 Plantilla de la Encuesta para la valoración de una voz normal:

I. DATOS GENERALES

Identificación: _____

FECHA: ___/_____/200_ Nombre de Evaluador: _____

- **Identificación**: Manera de identificar al hablante durante el proyecto y en la base de datos de señales de voz.
- **Fecha de realización:** Corresponde al día, mes y año, en que se lleva a cabo la evaluación.
- **Evaluador:** Corresponde al nombre de quién realiza la evaluación.

Nombre: _____Edad: ____años Sexo: F__ M__

Fecha de nacimiento: ___/____/____ Lugar de nacimiento: _____

Teléfono: _____ Dirección: _____

Ocupación actual: _____ Facultad: _____ Semestre: _____

Ocupación alterna y frecuente: _____

- **Nombre:** Corresponde al nombre completo del evaluado. Aunque no aparecerá en el proyecto será importante contar con esta información, en caso de necesitar verificar datos o ampliarlos.
- **Fecha de nacimiento:** Corresponde al año, mes y día del nacimiento del evaluado. Es importante contar con este dato para corroborar la edad del sujeto.
- **Edad:** Es el dato en años del evaluado. Esta información permitirá verificar si cumple con el criterio de inclusión del proyecto.
- **Teléfono:** Corresponde al número telefónico del evaluado, que permitirá mantener contacto con los integrantes de la muestra. Puede ser el fijo o el celular.
- **Dirección:** Corresponde a la ubicación en la ciudad del evaluado. Permitirá mantener contacto con los integrantes de la muestra.
- **Ocupación:** Corresponde a la actividad laboral que ejerce el evaluado. Permitirá verificar si cumple con el criterio de inclusión del proyecto. Docente, administrativo o estudiante, entre otros.

II. HÁBITOS

- ¿Hace uso frecuente de su voz? Si ___ No ___ ¿Cuanto tiempo en promedio?: ___horas/día

  Se refiere al tiempo durante el día que usa la voz como medio de comunicación. Si el evaluado habla 10 horas o más al día, puede ser considerado como un factor de riesgo para alteraciones de la voz, lo que implica que no puede formar parte de la muestra.

- ¿Practica oratoria, canto o una profesión que implique uso continuo de la voz? Si ___ No ___ ¿Cuál y cuanto tiempo en promedio?_____, ___horas/día.

  Al hacer esta pregunta se verifica el uso de la voz que realiza cada uno. En algunas ocasiones se puede ser estudiante pero también trabajar en una emisora como locutor, ser cantante aficionado, vendedor o tener otra actividad que implique el uso de la voz.

- Su voz es utilizada comúnmente a: Alta intensidad ____ Normal intensidad ____ Baja intensidad ___

  Permite verificar el volumen que utiliza el evaluado para comunicarse de manera habitual y la conciencia que tiene sobre este aspecto. Si la respuesta es alta indica que puede haber un sobre esfuerzo en el órgano fonador. Si es baja, pude indicar un problema de fuerza en el órgano fonador, dificultades auditivas o un hábito.

- Tiene cuidados con su voz: Hidratación: ___ Descansos cortos: ___ Descansos prolongadas: ___ otros: _____

  Se refiere a las estrategias que utiliza o no, el evaluado para cuidar el órgano fonador y la voz. Permite identificar los conocimientos sobre el tema y los hábitos que tiene el evaluado sobre este aspecto, para saber si puede o no, presentar algunas alteraciones de voz.

- ¿Realiza esfuerzo al hablar?: Si ___ No ___

  Se refiere a la sensación que tiene el evaluado de fuerza en la garganta o en la laringe para juntar las cuerdas vocales, lo cual puede mostrarse por la marcada evidencia de las venas en el cuello. Si la respuesta es si, puede indicar que el evaluado tiene dificultades a nivel del órgano fonador o un mal hábito al fonar.

- ¿Se cansa o fatiga al hablar?: Si ___ No ___

  Se refiere a la falta de aire durante la fonación como cuando se esta agitado al realizar algún ejercicio. También se hace evidente cuando al iniciar una frase no logra terminarla porque se acaba el aire o no hay la suficiente fuerza para juntar las cuerdas vocales. Si la respuesta es si, puede indicar que el evaluado tiene dificultades a nivel del órgano fonador o un mal hábito al fonar.

- ¿Tiene perdidas de la voz repentinas?: Si ___ No ___

  Se refiere a las ocasiones en las que sin una causa aparente (gripa, abuso vocal –gritar-, consumo de cigarrillo, tos, vómito, entre otras) se presenta la falta total o parcial de la voz.

- ¿Ante cambios bruscos de temperatura hay alteraciones de voz?: Si ___ No ___

  Hace referencia a cambios en las características de la voz, es decir, en el tono, si es más grave o más agudo de lo habitual; en la intensidad, si es más débil o más fuerte

de lo habitual; en el timbre, si es nasal, chillón, áspero, entre otros y por último, en la duración, por ejemplo que se queda sin voz al producir una frase larga.

- ¿Consume alimentos muy condimentados?: Si ___ No ___

  Se refiere al hábito que puede tener el evaluado por consumir comidas picantes con pimienta o ají. Este tipo de productos puede producir alteraciones digestivas como gastritis, acidez estomacal, úlceras gástricas, reflujo gastroesofágico, que ocasiona incremento en la producción de los ácidos del estomago y a veces provocar cambios en la mucosa del órgano fonador y en la voz. Es posible que la respuesta positiva a este hábito tenga relación con alguna patología que presente el evaluado.

- ¿Consume alimentos muy calientes o muy fríos?: Si ___ No ___

  El hábito de consumir alimentos a temperaturas extremas puede implicar cambios bruscos en la temperatura del cuerpo, alterando su buen funcionamiento, por ejemplo, al hablar muchas horas y tomar agua muy fría, cuando la musculatura del órgano fonador esta aún caliente. Es posible que la respuesta positiva a este hábito tenga relación con alguna patología que presente el evaluado.

- ¿Consume alcohol?: Si ___ No ___, ¿cuantas veces en promedio?:_____ veces/semana
- ¿Consume cigarrillo?: Si ___ No ___, ¿cuantas veces en promedio?:_____cigarrillos/día

  Estos hábitos producen cambios en la mucosa de las vías respiratoria superior y digestiva y pueden ser factores de riesgo para alteraciones anatómicas y funcionales del órgano fonador, que afectan la producción de la voz y las características de la misma.

- Tiene o ha tenido hábito de: Gritar frecuentemente ____ Cantar ___ Imitar voces____

  Permite reconocer los hábitos del sujeto y el uso de la voz. Si realiza algunas de estas actividades se podrá encontrar un sobre esfuerzo vocal importante y como consecuencia algunas alteraciones anatómicas y funcionales de su órgano fonador.

III. ANTECEDENTES

- Ha presentado sensaciones en su garganta de: Cuerpo extraño___ Prurito___ Resequedad___ Dolor___ Carraspeo ___ Tos ___ Disfonía ___ Ninguna ___

  ¿Con que frecuencia / hace cuanto tiempo?_____

Se refiere a algunos signos y síntomas que permiten identificar la presencia de alteraciones anatómicas y funcionales del órgano fonador. Cuerpo extraño hace referencia a la sensación de algún objeto en las cuerdas vocales o en la garganta. Prurito se refiere a la sensación de picazón en la garganta. Resequedad se refiere a la sensación de la garganta seca que requiere consumir frecuente líquido. Carraspeo se refiere al sonido de umh, umh, que se hace con las cuerdas vocales, generalmente porque se tiene la sensación de cuerpo extraño, o acumulación de secreciones en ellas.

- Recibe actualmente tratamientos médicos o especializados para el cuidado de la voz: Si ___ No ___
  ¿Cuales? _____

  Hace referencia a los tratamientos con medicamentos, quirúrgicos o de terapia que se reciben en el momento de la evaluación o anteriormente, que puedan indicar que el paciente tiene o ha tenido alguna alteración anatómica o funcional que afecta la producción de la voz y que lo imposibilita para pertenecer a la muestra.

- Recibe tratamiento farmacológico Si__ No___
  ¿Cuál?_____

  Esta información hace referencia a cualquier otro medicamento que consuma el evaluado y que pueda generar cambios en el órgano fonador y modificar las características de la voz, por ejemplo medicamentos para tratamiento hormonal o algunos medicamento que producen gastritis y esto altera la mucosa de la laringe y el estomago.

- Tiene dificultad al realizar movimiento de: lengua _____ labios _____ mejillas ___ Mandíbula___

  En este aparte se pregunta por la capacidad de realizar movimientos de manera adecuada para la producción de los sonidos del habla (consonantes y vocales). En muchas ocasiones el paciente es conciente de las dificultades de movimiento que presenta y que pueden afectar la producción de los sonidos y en algunos casos la resonancia de la voz (nasalización).

- ¿Hace sobreesfuerzo al pronunciar algunos fonemas y/o palabras?: Si ____ No ____
  Cuales: _____

  Hace referencia a la identificación por parte del evaluado (y del encuestador) de algún tipo de esfuerzo significativo en la producción de fonemas y/o palabras. El

evaluado y/o el encuestador podrán reconocer algún tipo de problema en el aspecto psicológico, anatómico o funcional y que afecten el pronunciamiento de fonemas y palabras, impidiéndole pertenecer a la muestra.

- ¿Tiene movimientos corporales asociados al hablar?: Si ___ No ___
  ¿Cuáles?: _____

  Se refiere a los movimientos corporales que acompañan al sujeto durante el habla y que pueden percibirse en el momento de la anamnesis. Se logra vislumbrar algún tipo de problema en la fonación y que requiere un esfuerzo corporal para poder llevar a cabo dicho proceso.

- Tiene o ha tenido diagnóstico de: Labio y/o paladar fisurado _____ Parálisis Facial ____ Nódulos y/o Pólipos Vocales ____ Parálisis Vocal ____ Renitis Alérgica ____ Asma___ Bronquitis ____ Respiración Oral ____ Laringitis___ Faringitis ____ Perdida Auditiva _____ RGE ____ Gastritis ____ Hernia Hiatal ___
  ¿Hace cuanto tiempo?_____

  Permite precisar, según el caso, problemas funcionales vocales, respiratorios y auditivos, entre otros, que puedan generar alteraciones anatómicas y funcionales del órgano fonador y modificar las características de la voz. Estos diagnósticos que el evaluado manifiesta están en relación directa o indirecta con otros síntomas y signos que presenta el evaluado y que han sido manifestados en el transcurso de la anamnesis.

4.1.1.2 Los hablantes: Población muestra.

La muestra corresponde a estudiantes de la comunidad universitaria de las facultades de Ingenierías, de la Universidad del Cauca, que fueron contactados personalmente, siendo invitados a participar de esta muestra, ver Anexo D.1 y aprobaron firmando voluntaria y desinteresadamente el *consentimiento informado,* en el cual al participante se le expresa que la información suministrada de su parte será efecto de un análisis exclusivamente con fines académico para el desarrollo de este proyecto, ver Anexo D.2, (un análisis sobre las señales vocálicas de su voz).

Con las anteriores condiciones fueron seleccionadas 30 evaluaciones como muestra de a una población de voz normal, de las cuales 24 individuos, todos hablantes de lengua española asistieron y calificaron como individuos aptos para realizar los procesos de grabación, es decir, el 80% de los evaluados; los individuos muestreados presentaron un acento medianamente neutral, es decir que no presentan un marcado acento en su voz, o con un acento particular correspondiente a la región del suroccidente colombiano.

De los 24 hablantes grabados, se tienen voces correspondientes a 7 mujeres adultas, lo que equivale al 29.2% de la población muestra y 17 hombres adultos, lo cual corresponde al 70.8% de la población muestra.

El rango de edad tomado para la valoración de los hablantes varía entre los 19 y 28 años, éste rango se toma por estar referenciado por Jackson-Minaldi [1], como un periodo de poca oscilación en cuanto a los cambios de voz de los hablantes, que para nuestro registro se expresó de acuerdo con la Tabla No 4.1.

Grupo de Hablantes	Número de hablantes (N)	Rango de edades	Promedio de edades	Desviación estándar
Mujeres Adultas	7	19-28	22,29	1,98
Hombres Adultos	17	21-25	23,29	2,54

Tabla 4.1 Edades de los hablantes

4.1.1.3 La grabación de la voz

En el proceso de grabación de las voces se contó con el apoyo del director de la emisora de la Universidad del Cauca, Diego Ignacio Torres C. y con el operario calificado Jorge Gonzáles, estas grabaciones se realizaron en el cuarto de sonido de las emisora de la Universidad, espacio que por su características propias no modifica el espectro de la voz y no agrega reverberación[17] a los registros de las señales a evaluar, requisitos indispensables para un óptimo registro de las voces.

[17] La reverberación es la persistencia de las sensaciones auditivas en un local después de la emisión de un sonido. Real academia de la lengua española, *http://www.rae.es*

Los registros se obtuvieron usando un micrófono Sony C-48, ver figura 4.1, el cual posee una respuesta en frecuencia de 30Hz a 16Khz y fue colocado aproximadamente 20 cm. de la boca para captar principalmente el sonido directo de la voz y minimizar la influencia de la sala, pero no demasiado cerca con el fin de evitar la saturación del micrófono [1]; para aislar el micrófono de ruidos eventuales como golpes y vibraciones se instaló sobre una suspensión elástica fija a un pie pesado.

Se usó un equipo de cómputo con una tarjeta de sonido profesional Marca Creative E-MU 1820, ver Fig. 4.2. El micrófono es conectado directamente a la tarjeta de sonido y mediante el Software Cool Edit Pro v1.2[18], es capturada la señal de voz a una frecuencia de muestreo de 44100 Hz, 16 bits y en Stereo.

Figura 4.1 Micrófono Sony c-48

Figura 4.2 Tarjeta de sonido Creative E-MU 1820.

[18] El Software Cool Edit Pro v1.2 se encuentra licenciado para la Universidad del Cauca.

Se realizaron dos grabaciones de cada individuo muestreado, pidiéndole que pronunciara de manera natural, aislada y sostenida las 5 vocales. A cada hablante, se le ofreció un material escrito donde se le explicaba el modo de realizar la fonación en el orden /a, o, u, e, i/ en una primera serie y el orden /i, e, u, o, a/ en una segunda serie; orden que tiene la finalidad de ecualizar la energía de la fonación, evitando registros con energía insuficiente para su medición [38].

Previamente a los hablantes se les realizaron preguntas sobre algún tipo de afección temporal que los estuviera afectando tal como gripa, disfonía, resequedad, entre otras para reconocer que el estado de voz no se hallaba alterado en ese preciso momento, se les explicó el procedimiento y el desarrollo de las grabaciones con el fin de evitar posibles errores en el desarrollo de la grabación de las señales de voz y necesariamente se hicieron otras repeticiones a petición del hablante o del operario para asegurar la calidad del registro tomado; en promedio el proceso de grabación tomó de 5 a 10 minutos por hablante.

4.1.1.4 Implementación de la base de datos

El resultado del proceso de grabación de voz, es un archivo de gran valor auditivo ya que se realizó bajo el mayor cuidado y fidelidad en la señal de voz, además a dicho registro se le realizó una segmentación para tener por separado tanto los registros de cada hablante como los sonidos de cada vocal pronunciada.

La segmentación se realiza con el mismo programa que permite la captura, el Cool Edit Pro V1.2, en la Figura 4.3 se observa el proceso de selección del inicio y fin de cada vocal y se agregaba como una pista o track diferente, posibilidad que ofrece el software, ver figura 4.4, descartando las pistas intermedias que son registros de silencios. Ya realizado el procedimiento en todo el archivo se procede a guardar cada pista como un único archivo y a una frecuencia de muestreo de 8000 Hz, 16 bits y en Mono.

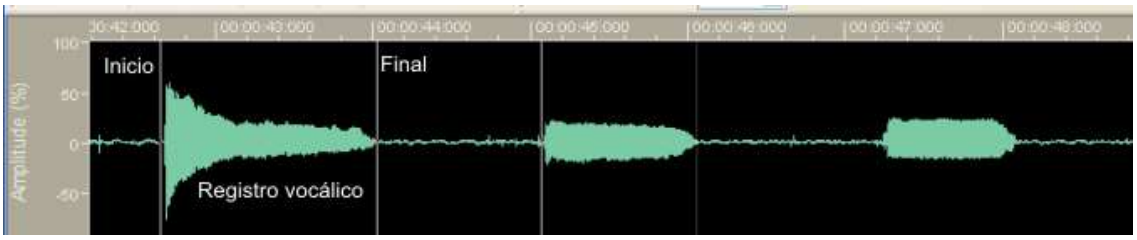


Figura 4.3 Selección del inicio y final de la vocal

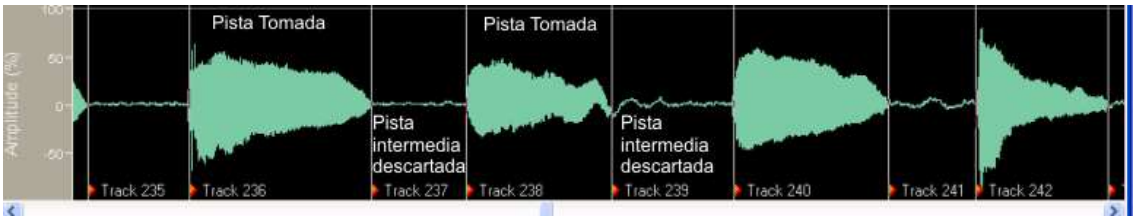


Figura 4.4 Pistas de las vocales divididas

Obtenido el registro digital de las señales vocálicas, se realiza el proceso de control y verificación para identificar registros defectuosos por causas de algún tipo de falla humana o técnica durante la grabación, o sobre posición de las señales debido al software de grabación.

4.1.1.5 Administración de la base de datos

Luego de realizar el proceso de control y verificación de las señales de voz, a cada hablante se le asignó una carpeta con su identificación que a su vez se clasificaron según el género en masculino y femenino y las cuales contienen dos subcarpetas correspondientes a la primera y segunda serie de grabación. Las carpetas clasificadas son organizadas de la siguiente manera:

- **Género (Masculino – Femenino)**: Son las carpetas Masculino y Femenino, que almacenarán las señales de voz de los individuos según su género.

- **Código (VocUDC*nn*):** Es el código asignado a cada hablante donde VocUDC hace referencia a la base de datos y nn es el número que se le asignó. Además

en esta carpeta se encuentra la carpeta Registro y un archivo de audio cuya información es el nombre del hablante.

- **Registro:** Es la carpeta que contiene los archivos con la información de la señal de voz registrada.

- **Señal Vocálica (a, e, i, o, u):** Es el archivo con la señal de voz en formato *wav* y se encuentra nombrada de acuerdo a la señal vocálica: /a e i o u/.

4.1.1.6 Formas de Ondas

En las gráficas se puede observar los fonemas vocálicos en el dominio del tiempo para las vocales /a/ /i/, con la voz de mujer *VocUDC20,* figura 4.5 y 4.6 y la voz de hombre *VocUDC15,* Figura 4.7 y 4.8 en un intervalo de 400 y 20 milisegundos.

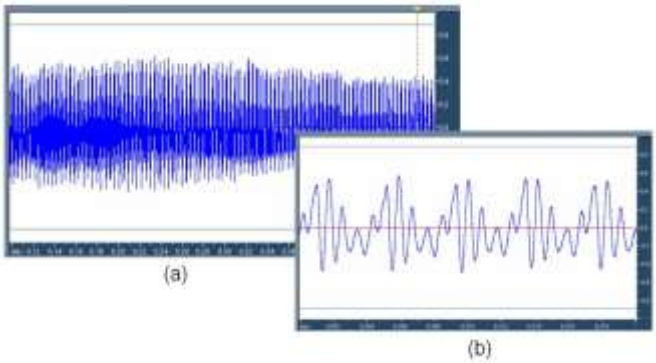


Figura 4.5 Vocal /a/ femenina: (a) intervalo de 400ms. (b) intervalo de 20 ms.

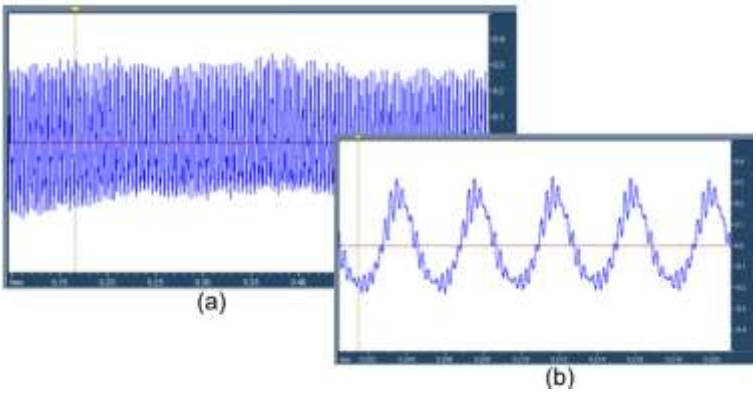


Figura 4.6 Vocal /i/ femenina (a) intervalo de 400ms. (b) intervalo de 20 ms.

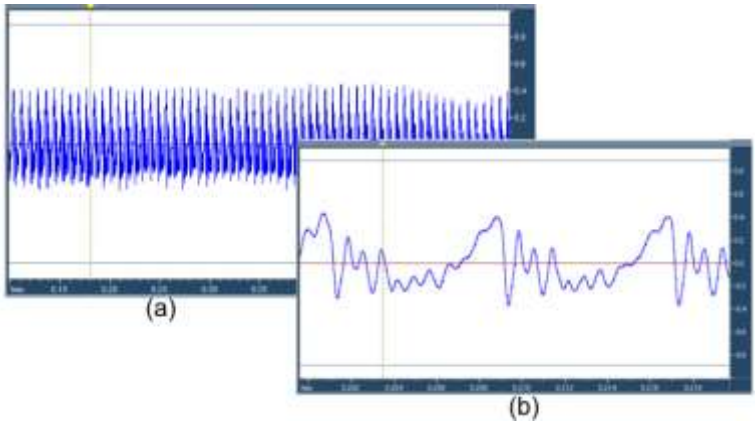


Figura 4.7 Vocal /a/ masculina (a) intervalo de 400ms. (b) intervalo de 20 ms.

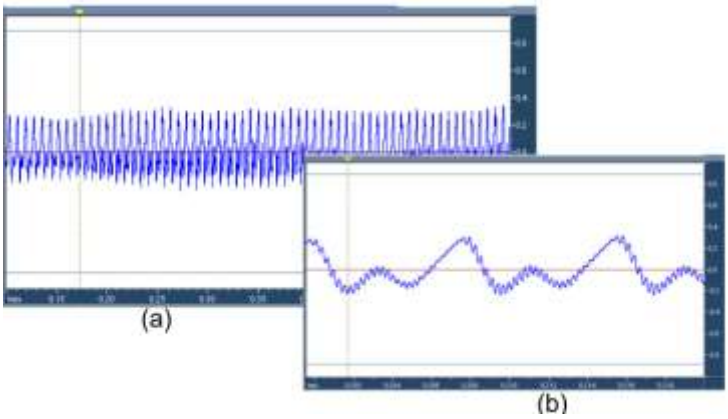


Figura 4.8 Vocal /i/ masculina (a) intervalo de 400ms. (b) intervalo de 20 ms.

Las otras formas de onda de todos los fonemas vocálicos *VocUDC20 y VocUDC15* en el dominio del tiempo se encuentran en el Anexo D.4

4.2 SPEECH FILING SYSTEM (SFS).

Para llevar a cabo la evaluación de nuestro sistema de estimación de la FF, se realizó una comparación entre el sistema desarrollado CGAWAVF, tratado en el capítulo 3, con otro sistema de calidad probada, dicho sistema es el software “Speech Filing System” – SFS, el cual ha sido usado en el tratamiento de señales de voz, como en la investigación de Jim Gilsinan IV [39].

4.2.1 Reseña del Speech Filing System

El “Speech Filing System” (SFS) tuvo su origen en el software desarrollado para la investigación del habla por la “University College London” (UCL), el “Imperial College London” y el GEC “Hirst Research Center”, todo esto bajo la iniciativa del proyecto software llamado SPAR; al finalizar este proyecto la UCL, retoma el código del SPAR y modifica sus librerías, de aquí nace el SFS, el cual ha sido continuamente desarrollado y usado en fonética y lingüística en la UCL desde 1987.

4.2.2 Estimación de la *FF* con SFS.

El procedimiento para la estimación de la *FF* con el SFS, empieza seleccionando el archivo de la señal vocálica que se encuentra en la base de datos VocUDC, desde la barra de Menú en la opción: *File* → *Open*, escogiendo la señal vocálica a tratar.

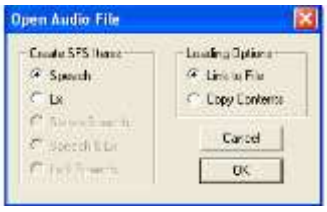


Figura 4.9 Archivo de Audio en SFS.

Al abrir una de las vocales del hablante se crea un ítem “*speech*” en el programa, con la opción para cargar el archivo *loading options*, como se muestra en la figura 4.9. De esta manera tenemos preparado el archivo de la señal vocálica y procedemos a utilizar la herramienta para la estimación de la frecuencia fundamental para la cual se selecciona la casilla *Speech* del ítem creado, seguidamente se ingresa en la barra de menú a la opción: *Speech* → *Analysis* → *Fundamental frecuency* → *Fundamental Frecuency Estimate* (ver figura 4.10), que ingresa un nuevo ítem llamado *Fx*, el cual en sus propiedades contiene los valores estimados de la FF.

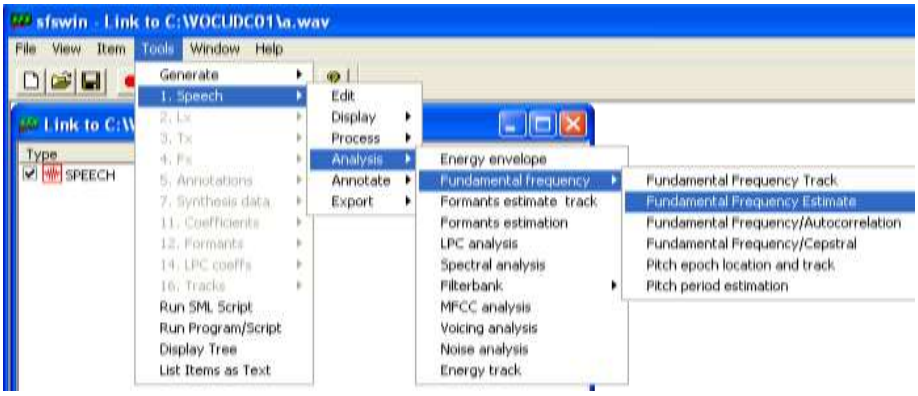


Figura 4.10 Estimación de la frecuencia fundamental con SFS.

Para observar las gráficas de la señal vocálica y los valores estimados es necesario seleccionar la casilla del ítem *Fx* y presionar el icono "*display checked ítems*" de la barra de herramientas. Para poder observar la gráfica de los espectrogramas de banda ancha se presiona el icono "*Wideband Spectrograms*" de la barra de herramientas de la última ventana emergente. En la figura 4.11 se observa la señal vocálica, seguida de los espectrogramas de banda ancha y su análisis que corresponde a la estimación de la FF, donde se muestra la frecuencia por cada segmento de la señal vocálica.

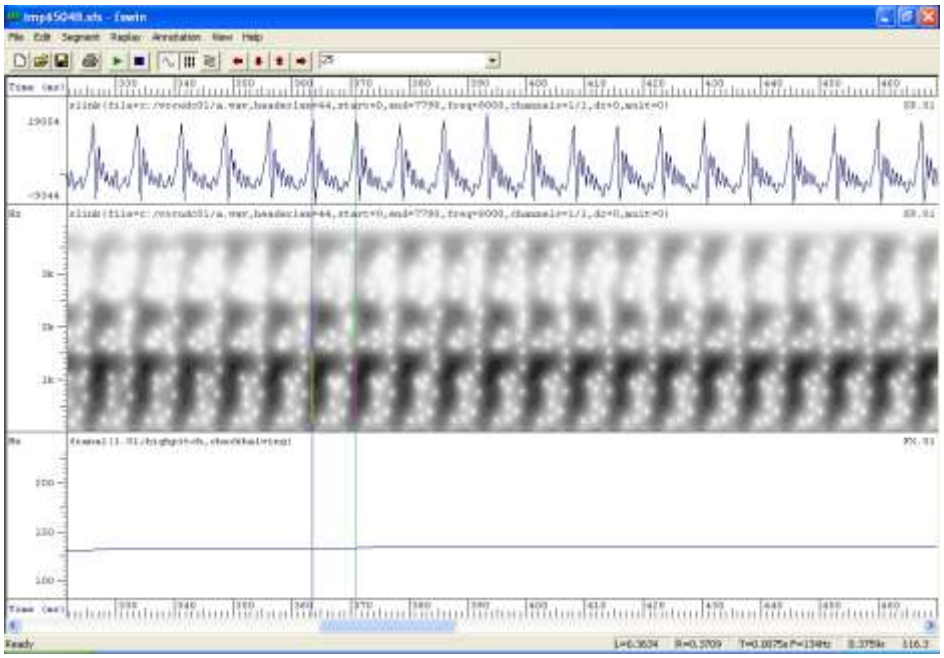


Figura 4.11 Ítems de la señal de voz y la estimación de la frecuencia fundamental

Al presionar el icono "*display propierties*" de la ventana principal se observa numéricamente los valores estimados de la FF representada en una matriz, figura 4.12.

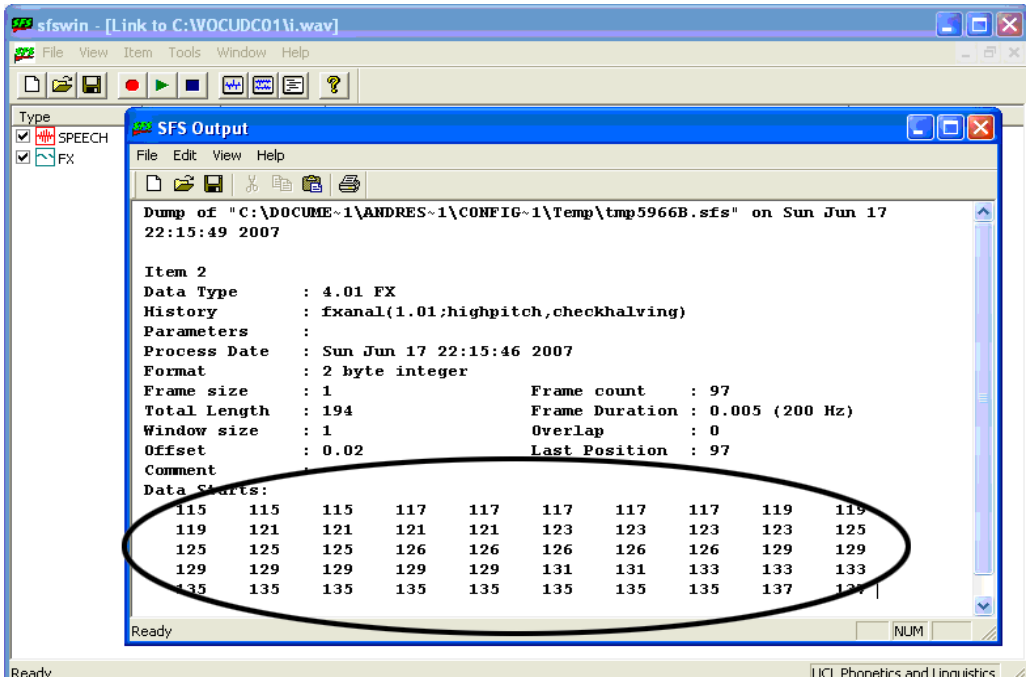


Figura 4.12 Matriz de valores estimados de la frecuencia fundamental

Seguido estos valores de la matriz son exportados a Matlab donde se le hará el tratamiento estadístico de Moda mediante el comando ***[M moda]=convertir([x]),*** donde ***x*** será la matriz resultante del SFS, este proceso se realizó con todos los registros de las vocales de los hablantes. El resultado dado es un valor estimado de la Frecuencia Fundamental del SFS. Los resultados de la estimación con el SFS, se encuentran en el Anexo D.5.

4.3 RESULTADOS.

El estudio en señales de voz femenina y masculina realizado bajo el sistema CGAWAVF, es un aporte académico importante ya que abre un espacio en el procesamiento digital de señales con la técnica *wavelet*, que tiene una gran aplicabilidad en varios campos de la medicina como en este caso especifico de la fonoaudiología y la otorrinolaringología, siempre y cuando se continué fortaleciendo esta área de investigación. Su profundización permitirá ahondar en el funcionamiento de la laringe gracias a la comparación de patrones normales y patológicos, la determinación de la

frecuencia de voz; y el diagnostico, control, tratamiento y monitoreo de los problemas fonatorios.

La ecuación 2.17 y 2.18 se adaptó con el parámetro $\frac{1}{\sqrt{\pi\sqrt{a}}}$ para que la wavelet gaussiana compleja tuviera correspondencia tanto en frecuencia como en magnitud con la wavelet gaussiana compleja de L. J. García [8].

Las bandas de análisis del sistema CGAWAVF utilizadas para hombres como en mujeres, son las comprendidas entre la 2 y la 9. Los resultados arrojados por el sistema en el procesamiento de señales de voz junto con los datos logrados a través del SFS, que se hallan representados en las tablas 4.2 y 4.3 y las figuras 4.13 y 4.14, muestran semejanza en la frecuencia fundamental estimada *FF* por ambos sistemas, tanto en hombres como en mujeres.

Sistema	Sexo	i [hertz]	e [hertz]	a [hertz]	o [hertz]	u [hertz]
SFS	**Fem.**	240	233	221	225	242
	Mas.	130	121	125	126	132
CGAWAVF	**Fem.**	239	236	226	237	248
	Mas.	140	134	137	134	142

Tabla 4.2 Valores promedios del SFS y CGAWAVF

Sistema	Sexo	i [hertz]	e [hertz]	a [hertz]	o [hertz]	u [hertz]
SFS	**Fem.**	22,49	23,46	33,27	12,69	17,93
	Mas.	14,16	18,32	13,54	11,56	12,15
CGAWAVF	**Fem.**	11,67	13,22	13,28	15,15	18,39
	Mas.	18,01	14,36	15,18	12,66	14,21

Tabla 4.3 Valores de la Desviación estándar del SFS y CGAWAVF

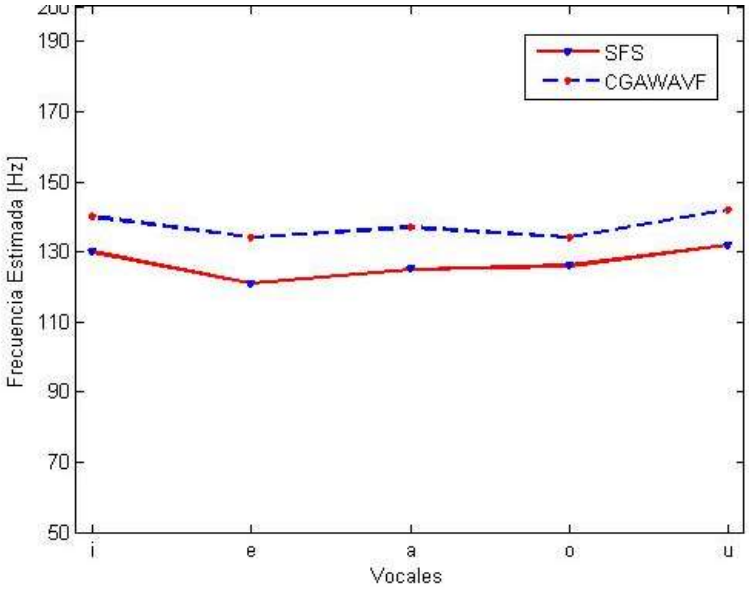


Figura 4.13 Valores Promedios en hombres SFS Vs CGAWAVF

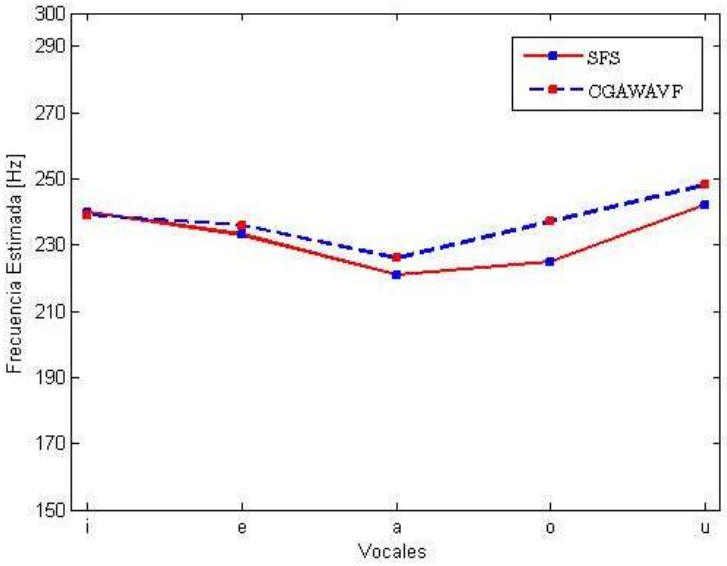


Figura 4.14 Valores Promedios en mujeres SFS Vs CGAWAVF

De igual forma, al comparar los datos experimentales con los referentes bibliográficos (Jackson-Minaldi [1]) se encuentra que existe una mínima diferencia en los valores de la frecuencia fundamental tanto en hombres como en mujeres. De esta manera, se afirma que los resultados del sistema CGAWAVF están en correspondencia con la teoría, que se ve reflejado en que los datos obtenidos se mantienen dentro de los valores de la frecuencia fundamental *FF* más frecuentes, que para mujeres corresponde a 250 Hz y en los hombres a 125 Hz.

Los resultados de las mediciones obtenidas por Aronson, Furmanski, Estienne y Rufiner [38] se observan en la tabla 4.4. Dichas mediciones realizadas para los hombres y mujeres distan un poco de las mediciones con el sistema CGAWAVF; hay una diferencia de 9 a 18 Hz para los hombres y de 21 a 44 Hz para las mujeres. Esta diferencia se debe a la influencia geográfica, fonatoria del acento y a que los registros de las señales corresponden a personas entre 19 y 35 años, hecho influyente debido a que la FF disminuye con la edad. Aunque se tratan de individuos de una región geográfica diferente a la del proyecto, los valores de la FF se encuentran dentro de los rangos establecidos según Jackson-Minaldi [1].

Sexo	**i [hertz]**	**e [hertz]**	**a [hertz]**	**o [hertz]**	**u [hertz]**
Fem.	207,00	205,00	205,00	204,00	204,00
Mas.	130,00	125,00	127,00	124,00	124,00

Tabla 4.4 Características acústicas de las vocales del español rioplatense

Las vocales del tipo [i-u] tienen una *FF* mayor que las de tipo [a-e], la diferencia va desde 4 a 25 Hz, es decir, la frecuencia por cada vocal en iguales condiciones deberían presentar una *FF* semejante; sin embargo, como se cita en [1], de acuerdo a Peterson, Lehiste y Peterson, Leíste, Black y Mohr las cuerdas vocales tienden a tener frecuencias fundamentales que son en promedio características, e intrínsecas a la voz misma ya que la *FF* varía continuamente según los patrones de entonación y acento.

5. CONCLUSIONES Y RECOMENDACIONES

Este capítulo contiene las consideraciones finales del proyecto; indicando en términos generales el desempeño del algoritmo desarrollado como una herramienta software de estimación de la frecuencia fundamental de señales de voz con *wavelets*. Se plantea además algunas sugerencias en cuanto a trabajos futuros orientados al mejoramiento del algoritmo, así como en el campo de investigación del procesamiento digital de señales unidimensionales con la teoría *wavelet* y sus aplicaciones.

5.1 CONCLUSIONES

Con el desarrollo del proyecto se ha diseñado e implementado un sistema de estimación de la frecuencia fundamental basado en la transformada *wavelet*, el cual proporciona un entorno de experimentación en el campo del procesamiento digital de señales de voz, que esencialmente ilustra una de las tantas aplicaciones de la teoría *wavelet.*

En cuanto al tipo de familias *wavelets* identificadas como las más adecuadas para el procesamiento de las señales de voz, se encuentran las familias Gaussiana, Daubeuchies, Spline y Morlet. Para este trabajo se seleccionó la familia *Gaussiana* que reúne la mayoría de las características de resolución tanto en tiempo como en frecuencia y tiene similitud con las señales de voz, permitiendo una estimación más certera de la frecuencia fundamental para los registros de voz de la base de datos VocUDC.

El algoritmo esta basado en un modelo auditivo real, del cual se han tomado 17 bandas que se comportan como filtros pasa banda distribuidos en el rango de frecuencias de 50 Hz a 3 Khz. De éstos, experimentalmente se concluye que de las bandas 2 a la 9 son las que mejor se comportan tanto con voces femeninas como masculinas, debido a que

las características espectrales de las wavelets de estas bandas están localizadas dentro del rango de valores de FF.

Por otro lado, con el propósito de evaluar los resultados proporcionados por el algoritmo, éstos se comparan con los obtenidos mediante el software SFS y con los valores habituales de frecuencia fundamental (Jackson-Minaldi [1]). El algoritmo basa su resultado en una metodología estadística, mientras el SFS entrega valores aproximados según la duración de la señal de voz. El algoritmo entrega un solo valor de frecuencia fundamental basado en la medida de tendencia central: moda; en cambio el SFS, basado en la técnica de autocorrelación, entrega una matriz de valores de frecuencia fundamental dentro de los cuales se encuentra el valor obtenido por el algoritmo, logrando de esta manera resultados satisfactorios al mantener medidas objetivas aceptables con respecto a los valores habituales y diferencias mínimas entre los valores dados por el algoritmo desarrollado y el SFS.

El algoritmo de estimación de la Frecuencia Fundamental de Señales de Voz del Sur Occidente Colombiano implementado, cuenta con las características mínimas que debe poseer cualquier sistema de estimación, que como se demuestra en el trabajo, no es cuestión sólo de aplicar una transformada sino que requiere la implementación de módulos estadísticos. No se puede desconocer además que el desarrollo del proyecto ha implicado un estudio y análisis a fondo tanto de la producción de la voz como de la teoría *wavelet* y su aplicación en tratamiento digital de éstas señales, principalmente, en el campo de la estimación de frecuencia fundamental.

Para la construcción de la base de datos de referencia VocUDC es necesario realizar el reconocimiento de una muestra en la población, encuestas, anamnesis del habla, pruebas piloto; a la vez tener características mínimas adecuadas para la realización de una grabación de voz limpia, con poca interferencia y que no modifique su espectro. De igual forma la información registrada en la anamnesis y la base de datos VocUDC, quedan como material muy valioso para los estudiantes y profesionales de fonoaudiología. De igual forma esto les puede permitir realizar un análisis más preciso y detallado del estado clínico de pacientes con malformaciones del aparato fonador. En consecuencia, el sistema implementado se puede utilizar como herramienta de apoyo para los profesionales de la salud.

En este proyecto se logró un trabajo interdisciplinario, asociado con el programa de Fonoaudiología de la Facultad de Ciencias de la Salud y la Emisora Radio Universidad del Cauca de la División de comunicaciones, que sirve de soporte matemático para nuevos trabajos.

5.2 RECOMENDACIONES Y LÍNEAS FUTURAS

Este proyecto de estudio de señales de voz con el sistema CGAWAVF, es un aporte académico importante en este caso especifico para la Fonoaudiología ya que abre un espacio en el procesamiento digital de señales con la técnica *wavelet* que permite un monitoreo para detectar el comportamiento (frecuencia, cualidad y regularidad) de las cuerdas vocales.

El tiempo para procesamiento, que para las 17 bandas es de un máximo de 35 segundos dependiendo del computador utilizado para el procesamiento, se puede disminuir realizando funciones más eficientes o implementándolas en lenguaje C. De igual manera la precisión de la frecuencia fundamental estimada, se puede mejorar adaptando el método estadístico hasta obtener el mejor resultado. Asimismo, teniendo como referencia mediciones con el laringografo, herramienta de difícil acceso en la ciudad, se puede obtener una evaluación más precisa del desempeño del algoritmo.

El proyecto tiene la posibilidad de extensiones significativas. Por ejemplo, este trabajo es base para realizar un modulo adicional de identificación de las vocales pronunciadas. Con los resultados de este trabajo de grado se establecen la bases para construir un sistema de identificación de hablantes dependiente del texto, ya no utilizando una sola vocal sino que se pueda hacer la identificación utilizando palabras. Por otro lado, aunque el sistema implementa una eliminación elemental de ruido, sin considerar la naturaleza del mismo, a partir de este modulo se puede realizar un proyecto que estudie la estimación de la frecuencia fundamental considerando señales de voz ruidosas teniendo en cuenta la naturaleza del ruido. Otra ampliación del proyecto es la escogencia automática de las bandas para determinado tipo de señales de voz, ya sean femeninas o masculinas. Finalmente, se puede agregar un modulo que permita la

segmentación automática de los registros voz que contienen todas las vocales. En cuanto a la base de datos VocUDC, se pueden adicionar grabaciones de voz con otros tipos de población muestra, diferentes en sus características y perfiles, ya sea de edad, profesión, entre otros.

De otro lado, uno de los campos más estudiados actualmente en cuanto a la aplicación de la teoría *wavelet* se refiere, es la implementación hardware de su transformada, implementando interfaces que interactúen con los equipos médicos que permita realizar diagnósticos más rápidos, certeros y objetivos; sin duda, es ésta un área propicia para la investigación y la experimentación que puede acarrear logros muy significativos.

Adicionalmente, el presente trabajo aporta un fundamento claro y firme alrededor de la estimación de la frecuencia fundamental de señales de voz con la técnica *wavelet* y aporta las bases para aplicaciones futuras en la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca. Es así como con este proyecto y gracias la amplia gama de aplicaciones de la teoría *wavelet* el reconocimiento de hablantes, los conversores de voz a texto, la validación de claves biométricas, la identificación de señales telefónicas, entre otros, serán la base para nuevos trabajos de grado y proyectos futuros.

BIBLIOGRAFÍA

[1] JACKSON-MENALDI María Cristina. LA VOZ NORMAL. Editorial Panamericana. Buenos Aires - Argentina. 1992. 233 p.

[2] LE HUCHE François, ALLALI André. LA VOZ. Anatomía y fisiología de los órganos de la voz y del habla. Tomo 1. Edición 2. MASSON S.A. Versión Española. Barcelona - España 1993.

[3] BROWARSKY David, MARTÍN Marcelo. USING A PC TO PERFORM REAL-TIME SIGNAL PROCESSING IN COCHLEAR IMPLANT RESEARCH. Tesis de grado Ingeniería en Biomédica. Uruguay. 2005

[4] MERLO G; FERNÁNDEZ V.; CARAM F.; PRIEGUE, R. y GARCÍA MARTÍNEZ, R. RECONOCIMIENTO DE LA VOZ MEDIANTE UNA RED NEURONAL DE KOHONEN. Universidad de Buenos Aires. Departamento De Informática, Facultad De Ciencias Exactas. Buenos Aires.

[5] Y. Shiu, C.-H. Yeh, and C.-C. J. Kuo, “Audio fingerprint extraction for content identification,” in *Proceedings of SPIE Internet Multimedia Management Systems IV, Vol. 5242*, Nov. 2003, pp. 55–64.

[6] RABINER, L.R., CHENG, M.J., ROSENBERG, A.E., and MCGONEGAL, C.A. (1976). A comparative performance study of several pitch detectors, *IEEE Transactions on Acoustics Speech and Signal Processing*, , 399-413.

[7] ZANUY F. Marcos, INFLUENCIA DE LA DETECCIÓN VOZ/SILENCIO EN RECONOCIMIENTO DE LOCUTOR , Escola Universitària Politècnica de Mataró, adscrita a la UPC. Barcelona.

[8] GARCÍA Leonard J. TRANSFORMADA *WAVELET* APLICADA A LA EXTRACCIÓN DE INFORMACION EN SEÑALES DE VOZ. Tesis Doctoral. Universidad Politécnica de Cataluña. Departamento de Teoría de la Señal y las Comunicaciones. Barcelona (España). Mayo 1998.

[9] CUESTA F. David. ESTUDIO DE MÉTODOS PARA PROCESAMIENTO Y AGRUPACIÓN DE SEÑALES ELECTROCARDIOGRAFICAS. Tesis Doctoral. Universidad Politécnica de Valencia. Departamento de Informática de Sistemas y Computadoras (DISCA). Valencia (España). Septiembre de 2001.

[10] Grupo de tratamiento avanzado de señales. ANÁLISIS LOCALIZADO DE LA SEÑAL DE VOZ EN EL DOMINIO DEL TIEMPO. Universidad de Cantabria.

[11] ORTEGA G. Javier, González R. Joaquín. ANÁLISIS LOCALIZADO DE VOZ. Universidad autónoma de Madrid. Madrid. España. Octubre de 2005.

[12] CUENE G. Hoover A. CARACTERÍSTICAS ACÚSTICAS DE LA VOZ DE LOS PROFESORES ADSCRITOS A LA ESCUELA DE REHABILITACIÓN HUMANA DE LA FACULTAD DE SALUD DE LA UNIVERSIDAD DEL VALLE. Tesis de grado. Programa académico de fonoaudiología. Universidad del Valle. Cali. Colombia. 2005.

[13] FAÚNDEZ Z. Marcos, FERNANDEZ B. Mónica. ESTUDIO DE LA INFLUENCIA DEL RUIDO Y DE LA VARIACIÓN TEMPORAL EN RECONOCIMIENTO DE LOCUTOR. XIII simposium nacional de la unión científica internacional de radio URSI'98, Pamplona p. 757-758, ISBN 84-89654-12-3. [Fecha de consulta Mayo 2006] Disponible en Internet <URL: http://eupmt.es/imesd/telematica/veu/ursi98.pdf>

[14] DURÁN U. Carlos A. ALGORITMO PARA LA DETECCIÓN N DE PITCH EN POLIFONÍA EN TIEMPO REAL. Tesis Magíster. Pontificia Universidad Católica De Chile. Escuela de Ingeniería. Santiago de Chile. Chile. Mayo 2004.

[15] SAN MARTÍN César, CARRILLO A Roberto. IMPLEMENTACIÓN DE UN RECONOCEDOR DE PALABRAS AISLADAS DEPENDIENTE DEL LOCUTOR. Revista facultad de ingeniería, u.t.a. (chile), vol. 12 n°1 2004, pp. 9-14. Chile

[16] LEMMETTY Sami. REVIEW OF SPEECH SYNTHESIS TECHNOLOGY. Master's Thesis. Department of Electrical and Communications Engineering. Helsinki University of Technology. Espoo. 1999.

[17] FAUNDEZ Pablo, FUENTES Álvaro. PROCESAMIENTO DIGITAL DE SEÑALES ACÚSTICAS UTILIZANDO *WAVELETS*. Tesis de Grado (Ingeniero Acústico). Universidad Austral de Chile. Facultad de Ingeniería. Valdivia (Chile). 2000.

[18] KASCHEL C. Hector, WATKINS Francisco. SAN JUAN U Enrique, COMPRESIÓN DE VOZ MEDIANTE TÉCNICAS DIGITALES PARA EL PROCESAMIENTO DE SEÑALES Y APLICACIÓN DE FORMATOS DE COMPRESIÓN DE IMÁGENES.1 Rev. Fac. Ing. - Univ. Tarapacá, vol. 13 Nº 3, 2005, pp. 4-10 Departamento de Ingeniería Eléctrica, Facultad de Ingeniería. Departamento de Tecnologías Industriales, Facultad Tecnológica. Universidad de Santiago de Chile.

[19] MALLAT Stéphane. A *WAVELET* TOUR OF SIGNAL PROCESSING. Academic Press. Second Edition. San Diego (California - USA).1999.

[20] AKANSU Ali N., HADDAD Richard A. MULTIRESOLUTION SIGNAL DECOMPOSITION. Transforms, Subbands, and *Wavelets*. Academia Press. Second Edition. San Diego (California - USA), 2001.

[21] ERELL Adoram, WEINTRAUB Mitchel. ESTIMATION OF NOISE-CORRUPTED SPEECH DFT-SPECTRUM USING THE PITCH PERIOD. IEEE transactions on speech and audio processing, vol. 2, No. 1, part i, january 1994.

[22] FFT TUTORIAL. University of Rhode Island Department of Electrical and Computer Engineering. Communication Systems.

[23] RUFINER, Hugo L., MILONE, H. Diego. SISTEMA DE RECONOCIMIENTO AUTOMÁTICO DEL HABLA. *Ciencia, Docencia y Tecnología,* mayo, año/vol. 2004- XV, número 028. Universidad Nacional de Entre Ríos Concepción del Uruguay, Argentina. pp. 151-177

[24] ÁLVAREZ M. Agustín. ALGORITMOS DE EXTRACCIÓN DE CARACTERÍSTICAS. Facultad de Informática, Universidad politécnica de Madrid.

[25] HERNANDEZ, D. Marianito. ANALISIS COMPARATIVO DE ALGORITMOS PARA REDUCCION DE RUIDO EN SEÑALES UTILIZANDO *WAVELET*S. Trabajo de Grado (Licenciado en Ingeniería en Electrónica y Comunicaciones). Universidad de las Américas. Escuela de Ingeniería. Departamento de Ingeniería Electrónica. Puebla, 2003. [Fecha de consulta septiembre 2006]. Disponible en Internet <URL: http://catarina.udlap.mx/u_dl_a/tales/documentos/lem/hernandez_d_m/portada.html>

[26] SAHA, S., IMAGE COMPRESION – FROM DCT TO WAVELETS: A REVIEW. ACM Crossroads. Student Magazine. [Fecha de consulta Jul 2006] Disponible en Internet <URL: http://www.acm.org/crossroads/xrds6-3/sahaimgcoding.html>

[27] WOLFRAMRESEARCH. wavelet Explorer Documentation. Coiflets. © 2007 Wolfram Research, Inc. [Fecha de consulta May 2006] Disponible en Internet <URL: http://documents.wolfram.co.jp/applications/*wavelet*/Fundamentalsof*wavelet*s/1.4.5.html >

[28] BIOGRAPHIES OF WOMEN MATHEMATICIANS. Ingrid Daubechies. Larry Riddle. Agnes Scott College. November 21, 2006. [Fecha de consulta Enero 2007]. Disponible en Internet <URL: http://www.agnesscott.edu/lriddle/WOMEN/alpha.htm>

[29] JARAMILLO G. Juan J., GARCIA G. Gustavo A. RECONOCIMIENTO DE HABLANTES USANDO TRANSFORMADA *WAVELET* Y DSP'S. Proyecto de Grado. Universidad del Quindío. 2003.

[30] SEPULVEDA Alexander F., CASTELLANOS German. ESTIMACION DE LA FRECUENCIA FUNDAMENTAL DE LAS SEÑALES DE VOZ USANDO TRANSFORMADA *WAVELET*. Scientia et Técnica. Año X. No. 24. Mayo 2004. UTP, p. 7.

[31] LONG C. J., DATTA S. WAVELET BASED FEATURE EXTRACTION FOR PHONEME RECOGNITION. Department of Electronic and Electrical Engineering. Loughborough University of Technology Loughborough. LE11 3TU, p. 264-267. UK. *ICSLP-1996*. [Fecha de consulta Nov 2005]. Disponible en Internet <URL: http://www.asel.udel.edu/icslp/cdrom/vol1/239/a239.pdf>.

[32] MISITI Michael, MISITI Yves, OPPENHEIM Georges, POGGI Jean-Michael. *WAVELET* TOOLBOX FOR USE WITH MATLAB. User's guide version 3. 2006. [Fecha de consulta Mar 2007]. Disponible en Internet <URL: http://www.mathworks.com/access/helpdesk/help/pdf_doc/*wavelet*/*wavelet*_ug.pdf>

[33] CABEZAS B. Yaciro, GUEVARA C. Jairo. ALGORITMO DE COMPRESION Y RECONSTRUCCION DE IMÁGENES FIJAS APLICANDO LA TEORíA *WAVELETS*. Trabajo de Grado (Ingenieros en Electrónica y Telecomunicaciones). Universidad del Cauca. Facultad de Ingeniería Electrónica y Telecomunicaciones. Departamento de Telecomunicaciones. Popayán, 2005.

[34] "Waveform Audio File Format, Multimedia Programming Interface and Data Specification v1.0", Issued by IBM & Microsoft, 1991. [Fecha de consulta Mar 2006] Disponible en Internet <URL: ftp://ftp.cwi.nl/pub/audio/RIFF-format>

[35] PROAKIS John G., MANOLAKIS Dimitris G. TRATAMIENTO DIGITAL DE SEÑALES. 3ra. Edición, Prentice Hall. Madrid (España). 1998.

[36] GUARIN S. Norberto. ESTADÍSTICA APLICADA 2. Etapas del Método Estadístico. Estadística Universidad Nacional de Colombia. 2002. [Fecha de consulta Abr 2006] Disponible en Internet <URL: http://tifon.unalmed.edu.co/~pagudel/2etapas.html>

[37] PORTILLA Chimal E. ESTADISTICA, Primer curso. Primera Edición. Editorial INTERAMERICANA. México 1980. ISBN 968-25-0666-2.

[38] ARONSON L., FURMANSKI H., ESTIENNE P. RUFINER L., CARACTERÍSTICAS ACÚSTICAS DE LA VOCALES DEL ESPAÑOL RIOPLATENSE. Departamento de Implante Coclear – Fundación Arauz, Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET y la Facultad de Bioingeniería. Universidad Nacional de Entre Ríos. Argentina. [Fecha de consulta Feb 2007]. Disponible en Internet <URL: http://www.sinfomed.org.ar/Mains/info.htm >

[39] GILSINAN IV Jim. Yǒng Jǐu Fā Yīn: A SIMPLE MANDARIN CHINESE TONE RECOGNIZER. Thesis Bachelor of Arts. Harvard College. Cambridge, Massachusetts. 2001.

[40] MERTINS Alfred. SIGNAL ANALISIS. *Wavelets*, Filtres Banks, Time-Frequency Transforms and Applications. John Wiley & Sons Ltd. New York (USA). 1999.

Agradecimientos.

Los autores agradecen a todas las personas que hicieron parte formal de este proyecto:

Al Departamento de Telecomunicaciones y al GNTT.

Al Departamento de Fonoaudiología de la Universidad del Cauca y a la Fonoaudióloga Miryam Adela Barreto.

A la división de Comunicaciones de la Universidad del Cauca y a su operador Jorge Gonzáles.

A los hablantes que dieron un poco de su tiempo para las evaluaciones y el registro de sus voces.

Gracias a todos por su valiosa colaboración**.**