

*Ministerio de Educación Superior*  
*Universidad Central “Marta Abreu” de Las Villas*  
*Facultad de Ingeniería Eléctrica*  
*Centro de Estudios de la Electrónica y la Tecnología*  
*de la Información*



*Trabajo de Diploma*

*“Desarrollo de un software para la  
comprobación del proceso de rehabilitación en  
las consultas de logopedia y foniatría.”*

*Autor: Yolanney Rey González Díaz.*

*Tutor: Ing. Héctor Arturo Kairuz Hernández-Díaz.*

*SANTA CLARA - 2011*  
*Año 53 de la Revolución*

*Ministerio de Educación Superior  
Universidad Central “Marta Abreu” de Las Villas*

*Facultad de Ingeniería Eléctrica  
Centro de Estudios de la Electrónica y la Tecnología de la  
Información*



## *Trabajo de Diploma*

*“Desarrollo de un software para la comprobación del  
proceso de rehabilitación en las consultas de logopedia y  
foniatría.”*

*Autor: Yolanny Rey González Díaz.*

*Tutor: Ing. Héctor Arturo Kairuz Hernández-Díaz.  
Prof. Instructor*

*Centro de Estudios de la Electrónica y la Tecnología de la Información  
Facultad de Ing. Eléctrica. UCLV.  
E-mail: [akairuz@uclv.edu.cu](mailto:akairuz@uclv.edu.cu)*

*Consultante: Lic. Armando E. Alvaré Jaramillo  
Prof. Auxiliar*

*Departamento de Biología  
Facultad de Biología. UCLV  
E-mail: [armandoaj@uclv.edu.cu](mailto:armandoaj@uclv.edu.cu)*

*SANTA CLARA - 2011  
Año 53 de la Revolución*



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

---

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

---

Firma del Tutor

---

Firma del Jefe de Departamento  
donde se defiende el trabajo

---

Firma del Responsable de  
Información Científico-Técnica

## PENSAMIENTO

El futuro de un individuo no se mide por las circunstancias ni la suerte que se tenga, sino por los sueños, la voluntad y la fuerza que se empeñe para crearlos y hacerlos realidad.

Yolanny Rey González Díaz.

## TAREA TÉCNICA

1. Estudio de las características de la señal de voz para voces normales y patológicas.
2. Estudio de medidas objetivas que cuantifiquen de manera real las señales de voz.
3. Selección de algoritmos para la obtención de parámetros de la señal de voz (F0 y formantes).
4. Desarrollo de un software de fácil manejo y utilización para la extracción de características del habla como la frecuencia fundamental y los formantes.
5. Realización de pruebas para validar el funcionamiento del software con voces normales y patológicas grabadas con anterioridad.
6. Indicar posibles mejoras a las técnicas seleccionadas en virtud de su desempeño.

---

Firma del autor

---

Firma del tutor

## RESUMEN

El habla constituye uno de los principales protagonistas en la integración y desempeño del individuo en la sociedad. Su correcto desarrollo desde edades tempranas, así como la adecuada detección de patologías son aspectos prioritarios en el campo de la medicina que estudia la voz. En la actualidad no existe una explotación certera de técnicas que arrojen medidas objetivas utilizables para estos fines, a pesar de que esta rama de la Ingeniería Biomédica consta de importantes investigaciones tanto en nuestro país como en el resto del mundo. En el presente trabajo se desarrolló un software para la caracterización de voces normales y patológicas orientado a la utilización en consultas de logopedia y foniatría. Permite la extracción de parámetros de la voz como formantes y frecuencia fundamental. Además es posible la confección de un triángulo vocálico derivado de la extracción de formantes de tres vocales específicas (/a/, /i/, /u/). Para la extracción de los formantes se utilizó un método basado en el análisis LPC mientras que en la obtención de la frecuencia fundamental se utilizó el análisis mediante Cepstrum Real. El software brinda además la posibilidad de grabar una señal de voz y reproducirla. Posee una interfaz amena y de fácil interpretación haciendo más sencilla la utilización de técnicas matemáticas del procesamiento de voz a usuarios del software que no estén familiarizados con las mismas.

**Palabras claves:** formante, frecuencia fundamental, Cepstrum, Predicción Lineal, software.

**TABLA DE CONTENIDO**

PENSAMIENTO.....	IV
TAREA TÉCNICA .....	V
RESUMEN.....	VI
INTRODUCCIÓN .....	8
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA .....	10
1.1.    Procesos fisiológicos que intervienen en la producción del habla. ....	10
1.2.    Modelo matemático de la producción del habla.....	12
1.3.    Variables características del habla. ....	13
1.3.1.    Frecuencia fundamental y período fundamental de la voz. ....	13
1.3.2.    Formantes. ....	14
1.4.    Medidas acústicas relacionadas con desórdenes del habla.....	16
1.4.1.    Necesidad de medidas objetivas.....	17
1.4.2.    Medidas objetivas a tener en cuenta.....	18
1.5.    Conclusiones parciales. ....	18
CAPÍTULO 2: MATERIALES Y MÉTODOS.....	19
2.1.    Análisis mediante Cepstrum.....	19
2.2.    Análisis mediante Predicción Lineal.....	22
2.3.    Obtención de los parámetros de la voz utilizando LabVIEW 9.0. ....	24
2.3.1.    Cálculo de la frecuencia fundamental mediante detección de picos cepstrales. ....	25
2.3.2.    Método de detección de formantes mediante la extracción de picos espectrales utilizando LPC. ....	30
2.3.3.    Construcción del Triángulo Vocálico y cálculo de su área. ....	33
2.4.    Adquisición y reproducción de las señales de voz. ....	35
2.5.    Conclusiones parciales. ....	35
CAPÍTULO 3: RESULTADOS Y DISCUSIÓN.....	36
3.1.    Software FORMANTE. ....	36
CONCLUSIONES Y RECOMENDACIONES .....	44
REFERENCIAS BIBLIOGRÁFICAS .....	45

## INTRODUCCIÓN

El habla es la forma natural de comunicación del hombre y constituye la principal diferencia entre la especie humana y el resto de los animales. El lenguaje per se, se ha convertido en un campo de gran cantidad de investigaciones para la medicina y las aplicaciones de la tecnología. Los conocimientos desarrollados acerca de la fisiología de la voz y su procesamiento en el campo de las ciencias de la vida, tienen gran utilidad en sistemas de diagnóstico y constituyen la base de numerosas terapias de rehabilitación [1], sin embargo, por constituir la voz una señal compleja de la cual pueden extraerse muchos parámetros con el objetivo de modelarla o caracterizarla, este tema ha sido tan explotado como desconocido en plenitud.

Existen numerosas aplicaciones en el área de los sistemas digitales de comunicación de la voz, incluyendo el reconocimiento de locutores o palabras [2] y la respuesta oral por parte de las computadoras [3]. En las décadas del 70 y 80 del siglo pasado comenzaron a distinguirse otras esferas de aplicación para el tratamiento digital de la voz y en particular la detección de variables que la caracterizan; como el período fundamental, la frecuencia fundamental y los formantes, [4] [5] [6]. Haciéndose más notables los esfuerzos realizados por la comunidad científica, específicamente de nuestro país, entre finales del siglo pasado y principios del presente siglo, [7] [8] [9].

El desarrollo alcanzado en la medicina moderna ha hecho que el personal médico necesite cada vez más de algoritmos y equipos especializados para el diagnóstico y tratamiento de diferentes enfermedades. Las técnicas automáticas que permiten el diagnóstico de enfermedades de la voz están siendo cada vez más utilizadas por el personal de salud y el producto de su utilización cuenta con el aval de la comunidad científica moderna. El uso de técnicas de procesamiento digital de señales en la caracterización de señales de voz permite obtener medidas objetivas de los diferentes rasgos que caracterizan a estas señales biológicas y reduce la subjetividad introducida por la apreciación humana [10] [11].

En las consultas de Logopedia y Foniatría las evaluaciones subjetivas han sido la herramienta principal para la clasificación de las distintas patologías [12], por lo que surgen dudas sobre la confiabilidad y la validez de las mismas, especialmente por el hecho de que el proceso terapéutico y su seguimiento se lleven a cabo por técnicos y personal médico con



escasos recursos especializados. El proceso de atención y seguimiento de las patologías de la voz tratado desde el punto de vista subjetivo también genera una escasa vinculación de los pacientes de forma activa al tratamiento. Esto provoca que el componente emocional que conlleva a una recuperación en ellos sea nulo y pueda generar falta de compromiso en el proceso.

La globalización del conocimiento, así como las políticas internacionales hacia Cuba, hacen difícil la obtención, por parte de nuestro país, de herramientas que posibiliten una evaluación tangible de los pacientes en las consultas de Logopedia y Foniatría.

La tarea planteada para la realización de este trabajo se enmarca en las investigaciones que lleva a cabo el Laboratorio de Procesamiento de Voz, perteneciente al Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI), con la finalidad de encontrar medidas que apoyen el diagnóstico y evaluación de los pacientes con dificultades en la comunicación oral, que posibiliten su aplicación a los procesos de rehabilitación que se realizan en los centros de atención especializados de salud en nuestro país. Basado en lo anteriormente expuesto el **objetivo general** de este trabajo es:

- Desarrollar un software basado en programación en LabVIEW para su utilización en la evaluación del desempeño de los pacientes en las consultas de logopedia y foniatría.

Y como **objetivos específicos** se plantearon:

- Seleccionar el algoritmo adecuado para la determinación del período fundamental.
- Seleccionar el algoritmo adecuado para la determinación de formantes.
- Diseñar y desarrollar un software para el registro de características de la voz.

## CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

### 1.1. Procesos fisiológicos que intervienen en la producción del habla.

El habla es producida por la excitación acústica de una cavidad que varía su posición con el tiempo: el tracto vocal, región de la cavidad vocal que va desde los labios hasta las cuerdas vocales, Figura 1.1 [1]. En el habla no solo interviene el aparato respiratorio, existen otras estructuras de naturaleza nerviosa, como los centros nerviosos específicos de control del habla situados en la corteza cerebral (Área de Brocca y Wernicke), centros de control respiratorio en el encéfalo (Región Pneumotáxica y aunéstica) y estructuras de articulación y resonancia de la boca (la lengua, los labios, y el paladar blando) y cavidades nasales [3].

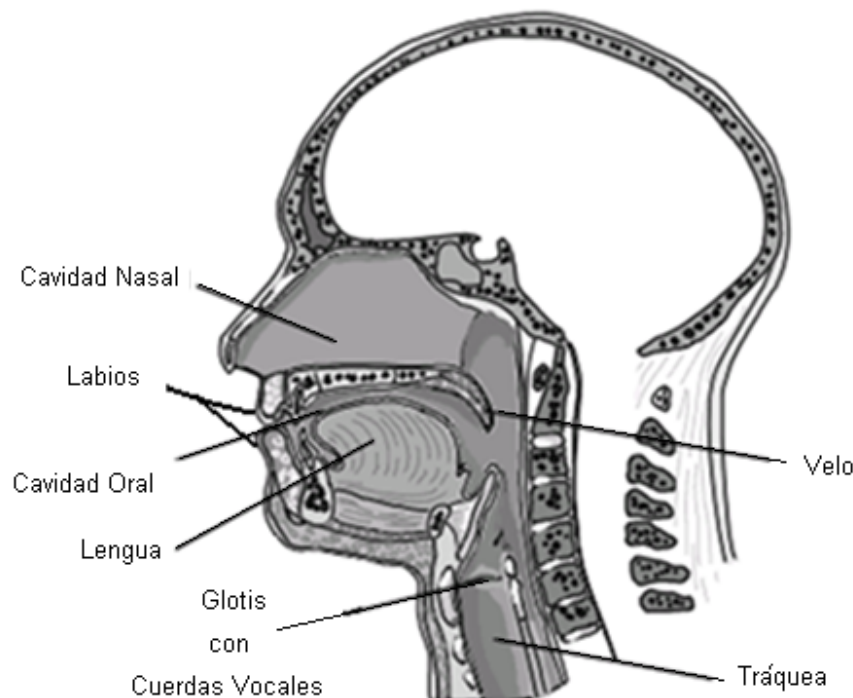


Figura 1.1: Representación del tracto vocal y estructuras anatómicas relacionadas con el habla, mediante un corte transversal del sistema fonador.

La producción del habla está compuesta por dos funciones mecánicas: la fonación y la articulación. En la fonación los pliegues de la glotis se cierran de tal forma que el paso del

aire por ellos produce que estos vibren. El tono de la vibración está determinado por el grado de tensión de los pliegues, la fuerza con la que se aproximan entre sí y por la masa de sus bordes. La articulación y la resonancia están determinadas por estructuras como la lengua, los labios, y el paladar blando mencionadas anteriormente.

Cuando el aire sale de los pulmones pasa a través de las cuerdas vocales, estas vibran y producen los sonidos del tipo sonoro; los que se diferencian entre sí por la variación de las características del tracto vocal y tienen una forma de onda cuasi periódica. Las vocales y los fonemas nasales son un ejemplo de algunos de estos sonidos. Los sonidos del tipo sordo, tiene una forma de onda básicamente aleatoria, con comportamiento similar al ruido, las cuerdas vocales se mantienen con una apertura constante y la forma del sonido varía según la posición del tracto vocal, dentro de esta clasificación se encuentran consonantes como la /s/ y la /f/.

La señal de la glotis se produce con características espectrales únicas, una serie de armónicos dependientes de una frecuencia específica (frecuencia fundamental), que al pasar por el tracto vocal, son modelados produciendo variaciones en el espectro de la señal, Figura 1.2. Como resultado de este fenómeno una serie de frecuencias son atenuadas, mientras que se realzan otras, estas últimas son conocidas como formantes y se identifican con números para su diferenciación (F1, F2, F3, etc.). La forma de la señal varía fundamentalmente según la posición de la lengua y la radiación labial.

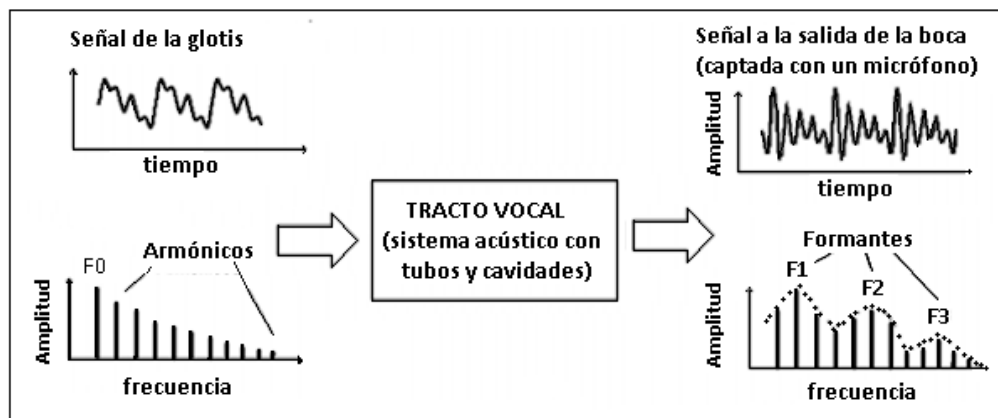


Figura 1.2. Esquema del efecto del tracto vocal sobre la señal de la glotis. A la izquierda la señal de la glotis en el dominio del tiempo y en el dominio de la frecuencia, a la derecha dicha señal modificada por el tracto vocal. (Tomado de [11]).

## 1.2. Modelo matemático de la producción del habla.

Producto a la complejidad del proceso de producción del habla, se han desarrollado modelos matemáticos para lograr una comprensión más sencilla del funcionamiento de este sistema [1] [13] [14]. Estos modelos tienen en cuenta las características de las estructuras más significativas que intervienen en la fonación y las funciones principales de las mismas. El modelo más utilizado es el llamado “Fuente-Filtro”, Figura 1.3, [14]. Este modelo puede ser dividido en dos partes: la primera consiste en dos fuentes de excitación, un tren de pulsos y ruido blanco gaussiano y la segunda en un filtro lineal que simula el tracto vocal. Se selecciona una de las excitaciones para la conformación de la señal, entre el tren de pulsos y el ruido blanco gaussiano, simulando la intervención de la glotis, de manera que si se está generando ruido gaussiano equivale a la producción de sonidos sordos, si se genera un tren de pulsos equivale a la producción de sonidos sonoros. La modificación producida por el resto de las partes del tracto vocal es simulada por el filtro lineal con función de transferencia  $H(z)$ , mostrada en la Ecuación 1.

$$H(z) = \frac{\sum_{i=0}^p a_i z^{-i}}{\sum_{i=0}^q b_i z^{-i}}, \quad 1$$

Donde  $a$  y  $b$  son los coeficientes del filtro,  $p$  y  $q$  su longitud.

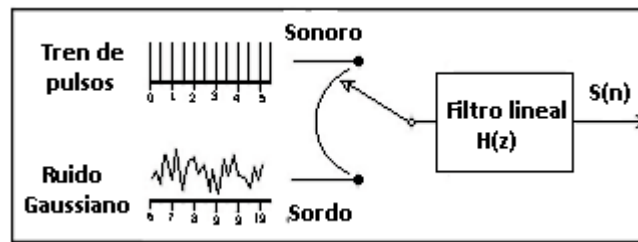


Figura 1.3: Representación del modelo de producción del habla “Fuente-Filtro”. La señal de salida  $s(n)$  es obtenida mediante la selección del tren de pulsos y el ruido gaussiano para obtener sonidos sordos o sonoros, y la modificación de estos por el filtro lineal.

La señal  $s(n)$  es producto de la modificación de la excitación seleccionada por el filtro lineal. Este modelo simula de manera matemática parte del proceso fisiológico relacionado con la producción de la voz.

### 1.3. Variables características del habla.

La voz es una onda compleja que, al constituir una señal biomédica, puede ser analizada, caracterizada, descrita y sintetizada con el objetivo de estudiar e investigar al ser humano. De ella pueden extraerse varios parámetros como la frecuencia fundamental ( $F_0$ ), el período fundamental ( $T_0$ ) y los formantes. Estos parámetros permiten modelar y caracterizar la voz y son necesarios para documentar el comportamiento del complejo aparato fonador y alteraciones en su desempeño en presencia de alguna patología.

#### 1.3.1. Frecuencia fundamental y período fundamental de la voz.

La frecuencia fundamental se determina por la velocidad a la cual las cuerdas vocales se abren o cierran durante la fonación de un sonido de tipo sonoro [15]. Mientras que el período fundamental puede definirse como: "el lapso de tiempo entre dos pulsos glotales sucesivos comenzando la medición en un punto arbitrario del ciclo glotal, Figura 1.4. El punto elegido está en dependencia del algoritmo individual escogido, pero para un mismo algoritmo se mantiene localizado en la misma posición" [7]. En la mayoría de la literatura técnica sobre procesamiento de voz, suele usarse el término *pitch* como un denominador común y nombre general para ambos términos ( $F_0$  y  $T_0$ ) [16].

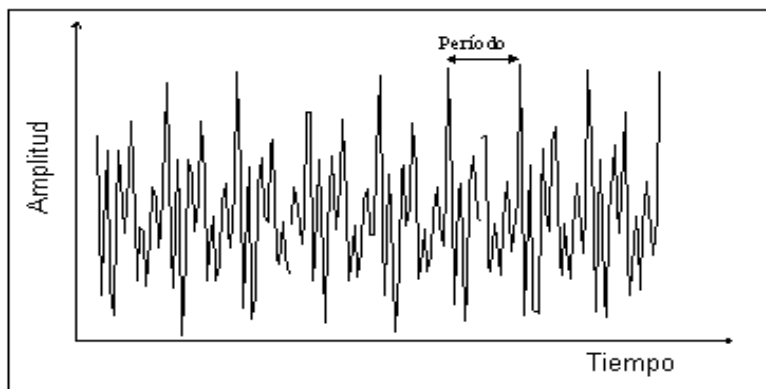


Figura 1.4: Representación del período fundamental de la voz, en un segmento de una señal del habla. (Tomado de [7]).

En el habla, las variaciones de la frecuencia fundamental contribuyen a la prosodia con diversas funciones, por ejemplo en la afirmación enfática, en los diferentes matices de un enunciado (afirmativo, interrogativo, etc.) y en las entonaciones regionales, entre otras. En

lenguas tonales como el Chino Mandarín, el contorno de la frecuencia fundamental con el que se pronuncia cada sílaba sirve para crear contrastes fonológicos [18].

El período fundamental de la voz contiene información fonética, prosódica y musical. Esto lo hace útil para los estudios lingüísticos y de fonética [9]. En la esfera de la medicina se aplica la detección del tono fundamental conjuntamente con otras tecnologías para dar apoyo al diagnóstico de enfermedades o al estudio de las mismas [17].

Entre la  $F_0$  y el  $T_0$  existe una estrecha relación, lo que permite el cálculo de uno a partir de la estimación del otro, como muestra la Ecuación 2, [7].

$$F_0 = 1/T_0 \quad , \quad 2$$

Sus valores difieren según el sexo, la edad y el estado de salud del individuo, aunque se ha demostrado que los valores normales de  $F_0$  para hombres y mujeres se encuentran alrededor de 100 y 200Hz, aproximadamente [18], siendo las frecuencias mayores características de las mujeres; mientras que las frecuencias pertenecientes a voces de niños resultan más agudas y están por encima de este rango.

### 1.3.2. Formantes.

Los formantes son las frecuencias de resonancia del tracto vocal y caracterizan las diferencias de los distintos sonidos. Se determinan por el proceso de filtrado por resonancia que se produce en el tracto vocal por la configuración de los articuladores. Técnicamente son bandas de frecuencia donde se concentra la mayor parte de la energía de un sonido.

La vibración de las cuerdas vocales produce ondas sonoras con un espectro de frecuencias bien distribuido, estas son filtradas por el tracto vocal y algunas frecuencias son reforzadas y otras atenuadas por el mismo. Las frecuencias fuertemente reforzadas son precisamente los formantes principales de la emisión sonora.

En muchas lenguas como el castellano, los dos primeros formantes son los que permiten distinguir la mayoría de sonidos vocálicos del habla. Típicamente el primer formante ( $F_1$ ), el de frecuencia más baja, está relacionado con la abertura de la vocal que en última instancia se vincula con la frecuencia de las ondas estacionarias que vibran verticalmente

en la cavidad. El segundo formante (F2) está relacionado con la vibración en la dirección horizontal y es respectivo a si la vocal es anterior, central o posterior.

Normalmente sólo F1 y F2 son necesarios para caracterizar una vocal, especialmente en las lenguas con menos de seis vocales Figura 1.5, [19]. En las vocales, los dos primeros formantes se determinan principalmente por la posición de la lengua. F1 tiene una frecuencia más alta cuanto más baja está la lengua, es decir, cuanto mayor abertura tenga una vocal, mayor es la frecuencia en que aparece el primer formante. Mientras que F2 tiene mayor frecuencia cuanto más hacia delante está posicionada la lengua, es decir, cuanto más anterior es una vocal, mayor es F2.

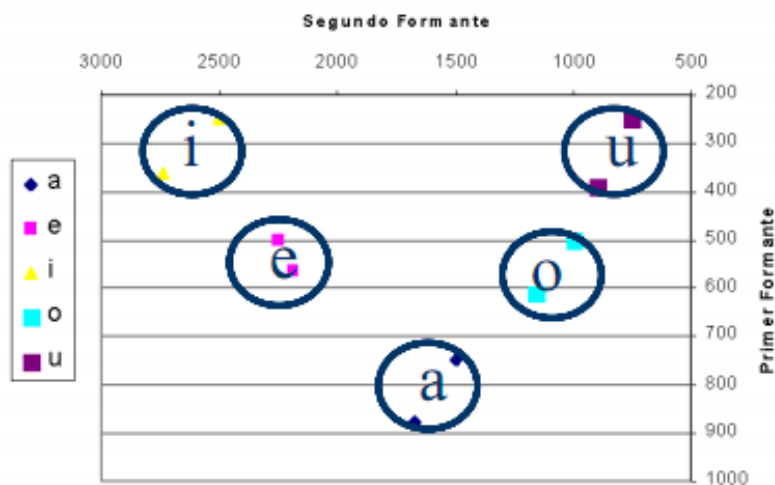


Figura 1.5. Representación del rango de valores normales de los formantes F1 y F2 para vocales en el habla española (Tomado de [11]).

Con los valores estimados de F1 y F2 para tres vocales específicas (/a/, /i/, /u/ generalmente) se puede construir un triángulo llamado Triángulo Vocálico (TV), Figura 1.6, ubicando en un sistema de coordenadas a F2 en el eje (y) y a F1 en el eje (x) para cada vocal. La forma del TV caracteriza la pronunciación de las vocales. De esta manera si el triángulo se encuentra reducido en tamaño se puede afirmar que hay poca diferencia entre la pronunciación de las vocales. Además los vértices del mismo pueden encontrarse también fuera de lugar para vocales pronunciadas de una forma incorrecta.

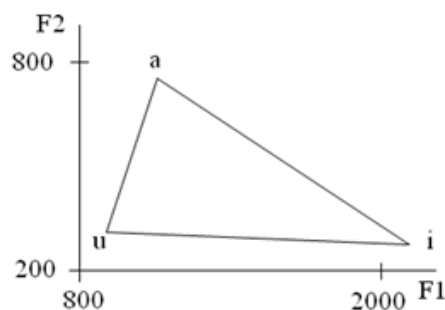


Figura 1.6. Representación de un ejemplo del Triángulo Vocálico para vocales (/a/, /i/, /u/) del castellano.

La estimación del  $T_0$ , la  $F_0$ ,  $F1$  y  $F2$  es importante en aplicaciones como la codificación de voz, reconocimiento del locutor y el desarrollo de sistemas de ayuda a discapacitados. Además, la  $F_0$  se emplea en la determinación de la entonación y las características emocionales de la voz. Los formantes permiten la discriminación entre sonidos. Así mismo, existen patologías relacionadas con desórdenes de la voz, en las que estas características se ven distorsionadas de una manera específica lo que permite realizar un diagnóstico o evaluar la evolución de un paciente en un proceso de rehabilitación [17] [8].

#### 1.4. Medidas acústicas relacionadas con desórdenes del habla.

Para producir una señal de voz estable y adecuada es necesaria la coordinación en el acoplamiento de los diferentes subsistemas involucrados en su producción. En ocasiones existen afecciones que provocan alteraciones en el control sobre movimientos musculares y por consiguiente en la producción del habla, las que se agrupan bajo el término de Trastornos Motores del Lenguaje (TML) y son consecuencia de una lesión en el Sistema Nervioso (SN). Se conocen dos clases de TML: la apraxia y la disartria del lenguaje.

Apraxia es sinónimo de disartria cortical y desintegración fonética, puede ocurrir aisladamente o asociada a la afasia motora o de Brocca. El control muscular está intacto, pero existe una pérdida del conocimiento de la articulación de la palabra. No existe distorsión del sonido del lenguaje, sino sustituciones fonéticas [10].

Disartria es cualquier combinación de trastornos de la respiración, fonación, articulación, resonancia o prosodia que puede ser causada por incoordinación muscular, debilidad muscular o alteración del tono muscular [10]. Es un término relacionado con un grupo de enfermedades neuronales que se diferencian entre sí por la localización de la lesión en el SN. Estudios clínicos realizados sobre disartrias han revelado que existe una relación



causa-efecto entre la localización de la lesión en el SN y el tipo de disartria exhibida. La localización del daño en el SN determina las manifestaciones y los síntomas producidos por la enfermedad y por consiguiente, las perturbaciones de la voz que se observan. Consecuentemente, se puede establecer una correspondencia entre las perturbaciones de la voz y la localización de la lesión en el SN [9] [20].

Entre los tipos de disartrias más comunes se encuentran: disartria flácida, espástica, atáxica, hipocinética (Enfermedad de Parkinson), hipercinética (Corea, Distonía, Temblor orgánico de la voz) y disartria mixta (Esclerosis Lateral Amiotrófica, Esclerosis Múltiple y Enfermedad de Wilson).

Existen otros trastornos del lenguaje que se califican según su origen nervioso y que están relacionados con afectaciones específicas en regiones cerebrales asociativas, estos se conocen como afasias (de Brocca, de Wernicke o de Conducción).

#### 1.4.1. Necesidad de medidas objetivas.

La difícil comprobación del proceso de la rehabilitación en pacientes con problemas del habla y la articulación de las palabras, constituye un fenómeno común en las consultas de Logopedia y Foniatría de los centros de salud en nuestro país. Las evaluaciones subjetivas han sido la herramienta principal para la clasificación, descripción y seguimiento de problemas del habla, por este motivo han surgido dudas sobre la confiabilidad y la validez de las evaluaciones subjetivas, especialmente por el hecho de que el proceso terapéutico y su seguimiento se lleven a cabo por técnicos y personal médico con escasos recursos especializados, esto hace que los mismos sean ineficientes y den poca información de la evolución de dichas patologías [21]. También se cuestiona cómo una evaluación perceptual puede discriminar entre dos problemas que ocurren simultáneamente en más de un componente del aparato fonador.

Por este motivo existe una necesidad importante de incorporar al proceso de evaluación de pacientes con problemas del habla algunas medidas objetivas que cuantifique de manera real el desempeño del individuo en un proceso de rehabilitación o de diagnóstico. Sin embargo, dado que la voz es el resultado de la interacción de múltiples procesos fisiológicos que no se pueden capturar con una sola técnica de medición, no se puede obviar la necesidad de las evaluaciones subjetivas, hay que lograr la mejor combinación

posible donde cada medida: subjetiva y objetiva, aporte lo mejor de sí a la evaluación de los procesos de diagnóstico y rehabilitación.

#### 1.4.2. Medidas objetivas a tener en cuenta.

Las medidas relacionadas con las frecuencias de los formantes (F1, F2), así como el valor de la frecuencia fundamental son representativas ya que estos varían con las características del tracto vocal, la edad y el sexo. Es necesario tener en cuenta los valores de F0, F1 y F2, así como las diferencias entre ellos: (F2-F1, F1-F0) para el logro de una buena caracterización de la señal de voz [20].

Estas medidas han probado su validez en diferentes aplicaciones como el Triángulo Vocálico que fue aplicado a pacientes con implante coclear para documentar su evolución [8]. Además se han utilizado en otras aplicaciones de interés [22]. En cuanto a las fluctuaciones de las frecuencias de los formantes es una de las anomalías más reportadas en producción de vocales en pacientes con problemas del habla, específicamente en presencia de disartrias, ejemplo de esto es la centralización de formantes y frecuencias anómalas para vocales altas y frontales.

Otras características a tener en cuenta son las estadísticas y contornos de F0, estas se relacionan con la fonación por lo que pueden caracterizar la calidad del habla [10].

Las medidas en segmentos largos son más prometedoras, (contornos de F0 y declinación de formantes). Similares medidas han sido usadas para demostrar la decadencia de los pacientes en estudios a largo plazo de Parkinson.

#### 1.5. Conclusiones parciales.

Luego de una amplia revisión de la bibliografía, que incluyó los procesos que intervienen en la producción del habla así como las medidas y características de algunas variables propias de ella, se demuestra la necesidad de una herramienta que analice el fenómeno del habla. Para la realización de esta herramienta se considera que la determinación de parámetros como los formantes F1 y F2, la frecuencia fundamental y la construcción del TV ofrecen un punto de partida ya que son suficientemente representativos y su validez ha sido probada en otros sistemas.

## CAPÍTULO 2: MATERIALES Y MÉTODOS.

Para la obtención de variables de las señales de voz que caractericen la fonación de un paciente y su comportamiento ante un proceso terapéutico, es necesaria la selección de algoritmos matemáticos que obtengan dichas características de manera acertada. Por este motivo a continuación se da una modesta descripción de algunas herramientas que serán utilizadas con este fin en la implementación del software.

### 2.1. Análisis mediante Cepstrum.

El habla puede ser representada como la salida de un sistema de filtrado lineal cuya excitación y respuesta del sistema varían lentamente con el tiempo, como se explicó en la Sección 1.2. Como resultado de esta representación la señal de voz se puede definir como la convolución de la excitación  $u(n)$  y la respuesta al impulso del tracto vocal  $h(n)$ , Ecuación 1; lo que equivale a una multiplicación de los espectros de las señales  $u(n)$  y  $h(n)$  en el dominio de la frecuencia como muestra la Ecuación 2 [23].

$$x(x) = u(x) * h(x) , \quad 1$$

$$X(w) = U(w)H(w), \quad 2$$

El proceso de formación del habla hace que sea complicada la separación matemática de la fuente de excitación,  $u(n)$  y la respuesta al impulso,  $h(n)$  a partir de la señal de voz resultante,  $x(n)$ . Esto es debido a que la manipulación algebraica de un proceso resultado de la multiplicación de dos fenómenos, es más compleja que la manipulación del mismo proceso si este es producto de una suma. El logaritmo tiene la propiedad de descomponer un producto en una suma como muestra la Ecuación 3 y resolver este problema.

$$\log(a \times b) = \log(a) + \log(b), \quad 3$$

De esta forma la separación de la señal de excitación y la respuesta al impulso del tracto vocal puede ser efectuada como muestra la Ecuación 4.

$$\log (X(w)) = \log (U(w)) + \log(H(w)), \quad 4$$

Sin embargo la Transformada de Fourier de  $x(n)$ ,  $(X(w))$ , es una señal compleja, por lo que es necesario definir el logaritmo complejo de la siguiente forma:

$$\log (X(ejw)) = \log |(X(ejw))| + \text{jarg}[X(ejw)], \quad 5$$

El cálculo del Cepstrum de una señal utiliza las propiedades del logaritmo para resolver el análisis de las características de una señal de voz, de manera más sencilla. Aplicando estas propiedades podemos llegar a que para señales de tiempo discreto, la definición del Cepstrum Complejo es la transformada inversa de Fourier (IDTFT) del logaritmo de la DTFT de una señal. De esta forma el Cepstrum de la señal  $x(n)$  se define como [23]:

$$c[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[X(e^{i\omega})] e^{i\omega n} d\omega, \quad 6$$

Donde  $X(e^{i\omega})$  es la DTFT de  $x(n)$  está definida por [24]:

$$X(e^{i\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n}, \quad 7$$

El Cepstrum Real  $cr(n)$  de una señal  $x(n)$  se determina como muestra la Figura 2.1, mediante el cálculo de la IDTFT del logaritmo de la magnitud de la DTFT, Ecuación 8:

$$cr[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log|X(e^{i\omega})| e^{i\omega n} d\omega, \quad 8$$

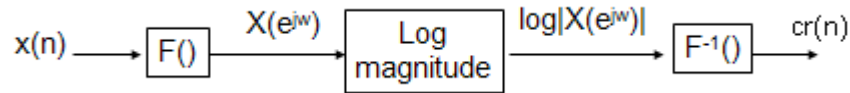


Figura 2.1. Diagrama en bloques de la obtención del Cepstrum Real  $cr(n)$  de la señal  $x(n)$ , a partir del cálculo de la Transformada Inversa de Fourier del logaritmo de la magnitud de la Transformada de  $x(n)$ .

Las bajas componentes cepstrales corresponden a variaciones lentas de las componentes espectrales y por tanto contienen información de la envolvente del espectro, la cual se relaciona con la respuesta en frecuencia del filtro que modela el tracto vocal. Estas componentes son usadas para aplicaciones relacionadas con el reconocimiento de voz [2].

En la Figura 2.2 puede observarse que el Cepstrum de una señal del habla contiene picos aislados en forma de impulsos a frecuencias específicas, que son múltiplos del período fundamental de la voz [25]. En cambio cuando el habla carece de actividad de voz no está presente periodicidad alguna producto a las características de ruido que exhibe este tipo de señal lo que provoca la ausencia de picos cepstrales [23].

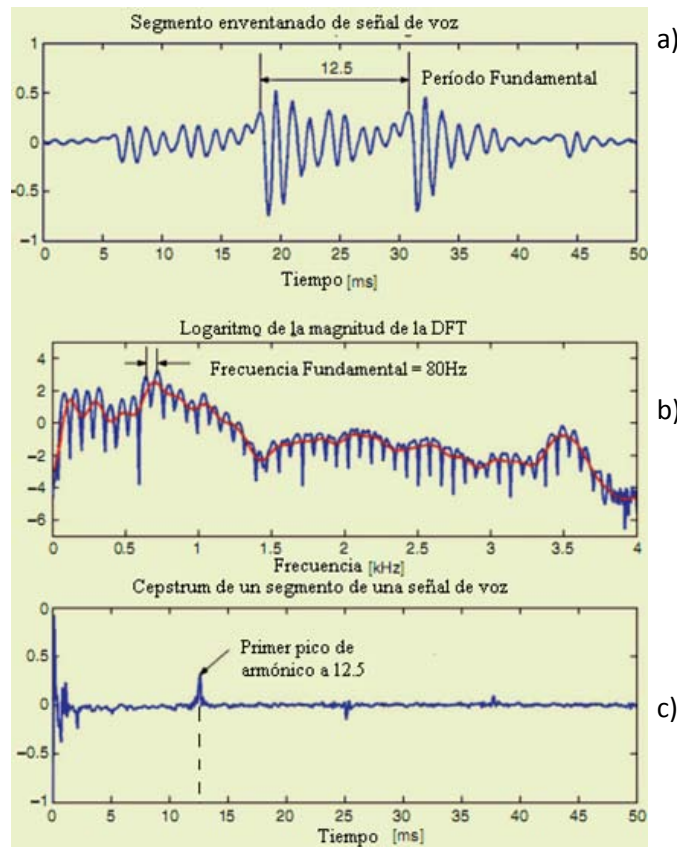


Figura 2.2. Representación de: **a)** segmento del habla de 50ms, enventanado utilizando una ventana Hamming, correspondiente a un sonido vocal con un período fundamental de 12.5ms aproximadamente. **b)**

logaritmo de la magnitud (en azul) y logaritmo de la magnitud, filtrado paso bajo (rojo) de la DFT del segmento de voz mostrado en a). c) Cepstrum del segmento de señal mostrada en a). (Tomado de [26]).

Como se muestra en la Figura 2.2, los picos encontrados en el Cepstrum de una señal de voz tienen una estrecha relación con el pitch de la misma [26]. La búsqueda de estos picos constituye la base para los algoritmos de detección de período fundamental de la voz mediante análisis cepstral como el descrito por “Noll” en [27].

Las facilidades matemáticas para el cálculo del Cepstrum, hacen que esta herramienta sea eficientemente utilizada en varios sectores como el procesamiento de la señal de voz, señales sísmicas, biomédicas y de sonar [26].

Además la fidelidad con que aparecen los picos cepstrales y su real correspondencia con el período fundamental de la voz contribuye al éxito del uso del Cepstrum en aplicaciones del procesamiento digital de voz [26] [27] [28] [29].

## 2.2. Análisis mediante Predicción Lineal.

La Predicción Lineal (LP por Linear Prediction) es una herramienta fundamental en diversas áreas como el filtrado adaptativo, estimación espectral y el procesamiento de voz [30]. La Predicción Lineal o Codificación mediante Predicción Lineal (LPC por Linear Prediction Coding) fue originalmente desarrollada por matemáticos y físicos que estudiaban análisis en series de tiempo en las décadas tempranas del siglo XX [13]. A medida que creció la importancia del procesamiento de voz, se fue centrando la atención hacia la aplicación de esta técnica en dicha rama.

En general el habla no puede considerarse como una señal determinística, ni estacionaria, sin embargo, si puede ser considerada como localmente estacionaria. Consecuentemente, LP puede ser aplicada en muchas áreas del procesamiento de voz tanto en reconocimiento, comprensión, caracterización, modelado, entre otras; logrando modelar de manera acertada el proceso de producción del habla [31].

La idea básica de LP es predecir muestras de una señal  $s(n)$  formando una combinación lineal de un número  $p$  de muestras previas. Usualmente la combinación es optimizada para minimizar el cuadrado del error de predicción, Ecuación 9, sobre un intervalo finito, entre las muestras actuales y las predichas. Determinándose un conjunto único de coeficientes de predicción ( $a_k$ ) [13].

$$e(n) = s(n) - \tilde{s}(n),$$

9

Donde  $e(n)$  es el error de predicción,  $s(n)$  es la señal de voz y  $\tilde{s}(n)$  es la señal predicha [2] [32].

La representación matemática para la predicción de  $s(n)$ , denotada  $\tilde{s}(n)$  puede expresarse como se observa en la Ecuación 10:

$$\tilde{s}(n) = \sum_{k=1}^p a_k s(n-k) \quad , \quad 10$$

Donde  $a_k$ ,  $1 \leq k \leq p$ , denota los coeficientes de predicción y  $p$  representa el número de coeficientes o el orden de predicción [13] [31]. Este modelo es el más utilizado, se aplica sobre muestras previas o salidas  $s(n)$  del sistema. Puede ser generalizado para incluir además las entradas,  $u(n)$ , del mismo. Si no se tienen en cuenta las entradas del sistema, el modelo se reduce a un modelo todo-polo o modelo Autorregresivo (AR) cuya ecuación en el dominio del tiempo es la Ecuación 10.

En el dominio de la frecuencia el análisis LP corresponde al modelado de la envolvente del espectro de la señal, lo que brinda una estructura espectral simplificada. Esto hace que esta técnica sea ampliamente utilizada en el cálculo de los formantes en el procesamiento de la señal de voz. En la Figura 2.3 se muestran dos espectros de una señal de voz, que corresponden a dos técnicas de análisis espectral, en el caso de a) mediante la Transformada de Fourier y b) mediante LP.

Entre ambas técnicas se evidencia una diferencia significativa en cuanto a la forma de los picos espectrales. Como la Predicción Lineal permite la obtención de un espectro más suavizado se pueden distinguir de manera más eficiente los picos que están presentes en el mismo. Esta característica hace que las técnicas que utilizan análisis de señales mediante LP sean altamente utilizadas en aplicaciones como la obtención de formantes de la señales de voz.

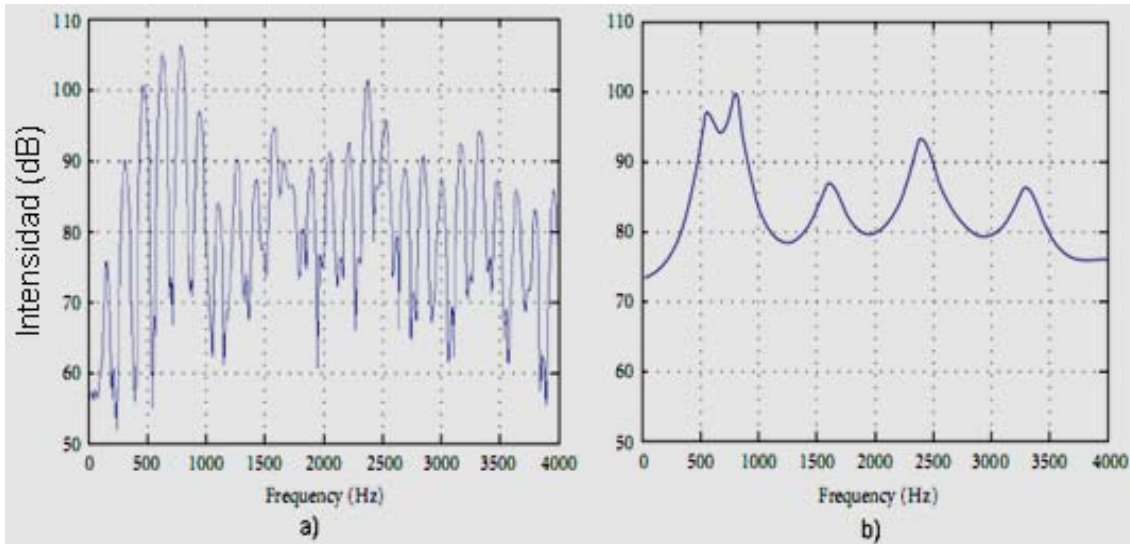


Figura 2.3. Representación de dos espectros de amplitud de una señal. a) Mediante la DTFT y b) mediante LP.

### 2.3. Obtención de los parámetros de la voz utilizando LabVIEW 9.0.

Con la utilización de los componentes básicos de la computadora, como la placa de audio comercial, es posible la utilización de dispositivos de sonido como micrófonos, bocinas o audífonos para la adquisición y reproducción de las señales de voz. El software LabVIEW 9.0 [33], contiene los drivers necesarios para manejar los datos provenientes de la placa de audio, además de brindar opciones como la de mostrar la forma de onda de la señal y graficar su espectro de frecuencias. Este software ofrece una plataforma de programación gráfica que resulta amena para la implementación de algoritmos matemáticos de procesamiento que posibiliten la extracción de variables características de las señales de voz. Además contiene una caja de herramientas de Procesamiento Avanzado de Señales que facilita la implementación de algoritmos matemáticos de alto nivel de complejidad. LabVIEW 9.0 ofrece la posibilidad de crear un instalador, un ejecutable y de incorporar un archivo de ayuda a los mismos que facilite la utilización del software. Igualmente el programa consta de una ayuda detallada y un set de ejemplos que constituyen un apoyo importante en la programación.

Por estas razones se utilizó la herramienta computacional LabVIEW 9.0 para la confección de un software que puede ser utilizado en los procesos de diagnóstico y rehabilitación de las consultas de logopedia y foniatría.



2.3.1. Cálculo de la frecuencia fundamental mediante detección de picos cepstrales. Existen distintos métodos para la estimación del período o la frecuencia fundamental de una señal cuasi periódica [34]. Si una señal es completamente estacionaria y periódica todas las estrategias utilizadas arrojan resultados idénticos. Sin embargo como la señal del habla es una señal no estacionaria, variante en el tiempo, aspectos de cada forma de medición como el punto de comienzo de la medición, el largo de los intervalos de medición, la manera de realizar el promediado, o el dominio de operación (tiempo, frecuencia, etc.), comienza a influir en el resultado y estos pueden diferenciarse entre algoritmos aunque cada uno sea correcto y preciso [16].

A partir del cálculo del período fundamental se puede determinar la frecuencia fundamental como el inverso del mismo, como muestra la Ecuación 2.

Los algoritmos a usar para el cálculo del pitch se basan en las características de la señal, tanto en el dominio de la frecuencia como el dominio temporal [15] [7] [35] [18]. En este trabajo se aplica un algoritmo basado en la detección de picos espectrales a partir de la obtención del Cepstrum como muestra la Figura 2.4.

El algoritmo utilizado se obtuvo a partir del artículo Using LabVIEW for Voice Signal Analysis, encontrado en la página web NI Developer Zone, de la National Instrument [36].

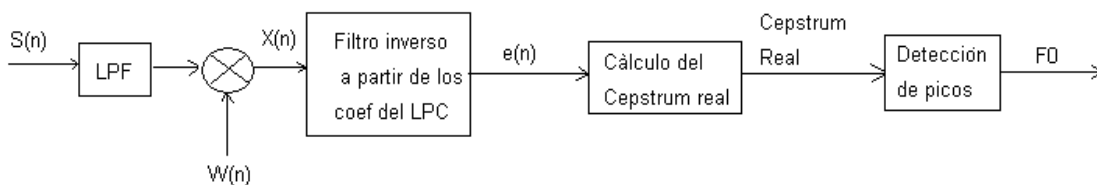


Figura 2.4. Diagrama en bloques del algoritmo de detección de la frecuencia fundamental utilizando detección de picos a partir de análisis cepstral.

La señal  $s(n)$  es filtrada paso bajo (LPF por sus siglas en inglés) y posteriormente dividida en tramas y enventanada mediante la ventana  $w(n)$  obteniéndose como resultado  $x(n)$ . Cada bloque de la señal es analizado para obtener los coeficientes de un modelo todo-polo del tracto vocal, los que son utilizados para realizar un filtrado inverso de las tramas de la señal, obteniéndose una señal residuo  $e(n)$ , a la que se le calcula el Cepstrum Real, que brindará los valores estimados de la  $F_0$  mediante una etapa de detección de picos [36].

Filtrado paso bajo de la señal  $s(n)$ .

La señal  $s(n)$  es filtrada para eliminar la interferencia de componentes de alta frecuencia, como los formantes, en el proceso de obtención de la frecuencia fundamental. En este caso se utilizó un filtro paso bajo con aproximante de Butterworth, con una frecuencia de corte de 1000Hz.

Enventanado de la señal.

La primera etapa de la detección del pitch, o en general de cualquier algoritmo sobre el procesamiento de señales de voz, es la segmentación de la misma en tramas y el enventanado de éstas. Este proceso es necesario debido a las características no estacionarias de las señales de voz, al tomar tramas cortas de esta señal garantizamos que sea localmente estacionaria y periódica en caso de que exista periodicidad en la misma.

Cuando se segmenta la señal usualmente las tramas son multiplicadas por una ventana temporal, para minimizar los efectos que pueden provocar los bordes resultantes de la segmentación. En la Figura 2.5 se pueden ver cuatro ejemplos de ventanas, son funciones que atenúan los bordes de los segmentos y mantienen prácticamente invariante la parte central del mismo.

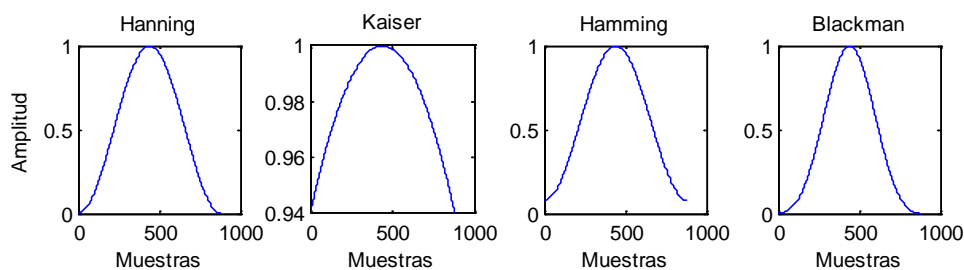


Figura 2.5. Representación de cuatro ventanas, Hamming, Hanning, Kaiser y Blackman, de una longitud de 40ms.

Para la selección de la ventana a utilizar se tuvieron en cuenta las modificaciones que hacen las mismas sobre la forma de onda de la señal. De acuerdo a la forma de la ventana estas pueden modificar de una forma o de otra la señal, en este caso se busca una ventana cuya supresión de los extremos de la señal no sea severa y que mantenga el valor de la señal en la parte central de la trama, buscando el cumplimiento de estas condiciones se evita la pérdida significativa de información de la señal.

En la Figura 2.6 se puede ver el efecto de cuatro tipos de ventanas sobre un segmento de una señal de voz. Se consideró el análisis de estas cuatro pues las ventanas Hanning,

Blackman y Hamming son de las más conocidas y utilizadas en el procesamiento de señales de voz, y al comparar su comportamiento con otro tipo de ventanas como la Kaiser se puede afirmar que su uso es más acertado en estas aplicaciones. En la Figura 2.6 se muestra que no existe una diferencia significativa entre las ventanas Hanning y Hamming. Siendo ambas las que mejores resultados alcanzan con respecto a las demás. Se seleccionó la ventana Hanning para su utilización en este caso, puesto que además de los resultados mostrados en la Figura 2.6, esta es utilizada con éxito en bibliografía consultada en este trabajo [37] [38].

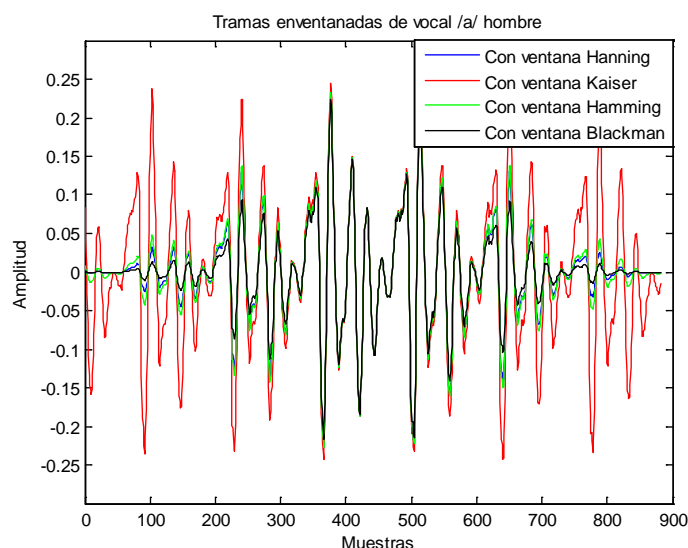


Figura 2.6. Trama de 40mseg de una señal /a/ de hombre, enventanada con cuatro tipos distintos de ventanas: Hanning, Kaiser, Hamming, Blackman.

Generalmente en algoritmos de procesamiento de voz de tiempo corto, se selecciona el tamaño de la trama de forma tal que sea lo suficientemente pequeña para que el parámetro a medir se mantenga constante dentro de la trama y lo suficientemente grande para que el mismo sea medible, en el caso de aplicaciones para la detección de F0 en señales de voz debe garantizarse que en la trama de señal se incluyan como mínimo dos o tres períodos completos [16].

En la mayoría de las aplicaciones de procesamiento de voz se utilizan tramas de 20 a 30mseg [15]. Sin embargo, cuando se aplica una ventana como se muestra en la Figura 2.7.b no hay un número suficiente de períodos de la señal dentro de la trama, esto disminuiría la confiabilidad del resultado de la extracción del pitch de la señal. Por este

motivo se determinó la utilización de una ventana de 40mseg, que ofrece una trama con cuatro períodos de la señal. Este tamaño de ventana equivaldría a 882 muestras si se muestrea la señal con una frecuencia de 22050Hz. En la Figura 2.7.a se puede observar como una trama de esta longitud logra contener una cantidad aceptable de períodos de la señal, sin que la trama llegue a considerarse excesivamente larga, este tamaño de ventana también se ha utilizado con éxito en otras ocasiones para la extracción del período fundamental de la voz [27].

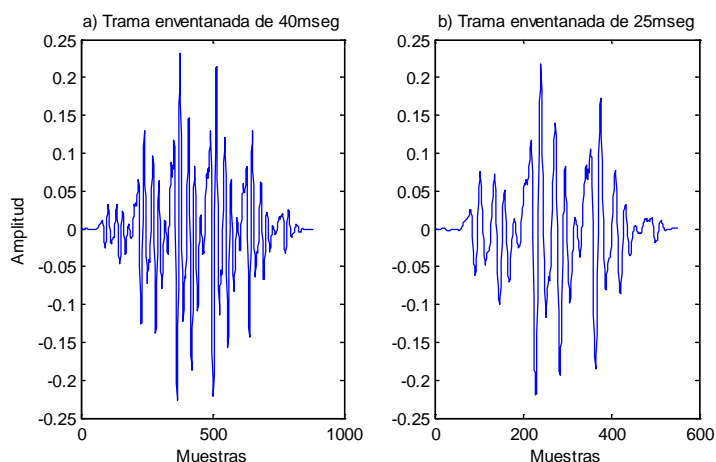


Figura 2.7. Tramas enventanadas de una señal de una vocal /a/ de hombre muestreada a 22050Hz, a) con un ventana Hanning de 40mseg y b) con un ventana Hanning de 25mseg.

Teniendo en cuenta lo mostrado en las Figuras 2.6 y 2.7 se determinó la utilización de una ventana Hanning de un tamaño de 40mseg para la segmentación de las señales de voz en el proceso de obtención de la frecuencia fundamental.

Filtrado inverso a partir de los coeficientes del tracto vocal.

La teoría del modelo fuente-filtro de la producción del habla provee una base para aplicar las técnicas de filtrado inverso. Este método se utiliza con el objetivo de separar la señal fuente del efecto del tracto vocal sobre la misma. Consecuentemente el resultado del filtrado inverso se corresponde a una estimación de la señal de la glotis [39].

La aplicación del filtrado inverso constituye una ventaja en la estimación de la frecuencia fundamental de las señales de voz, pues hace a un lado el efecto del tracto vocal eliminando componentes de la señal que inciden sobre la extracción del pitch.

En este caso el proceso de filtrado inverso se hace realizándole un filtrado FIR a la señal enventanada  $X(n)$ , a partir de los coeficientes LPC como coeficientes del filtro. Se utilizó la herramienta FIR filter.vi de LabVIEW 9.0 para el filtrado, la misma se muestra en la Figura 2.8.

De esta forma, como resultado del proceso de filtrado de  $X(n)$ , se obtiene la señal  $e(n)$ , que constituye una aproximación de la señal de la glotis. A partir de la señal  $e(n)$  se continúa el procesamiento para la extracción de la frecuencia fundamental de la señal de voz, según el diagrama en bloques de la Figura 2.4.

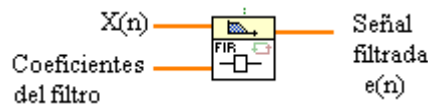


Figura 2.8. FIR filter.vi utilizado para la realización del filtrado inverso de la señal  $X(n)$ .

#### Obtención del Cepstrum Real.

El Cepstrum Real fue explicado en la Sección 2.1 y en este caso es obtenido utilizando TSA Real Cepstrum.vi, perteneciente a la caja de herramientas de Procesamiento Avanzado de Señales de LabVIEW 9.0 [33], mostrado en la Figura 2.9. Donde se obtiene a la salida el Cepstrum de la señal  $e(n)$  mediante el método definido en *método*, en este caso se utilizó Transformada Rápida de Fourier [24] y el tipo de ventana (Hanning) se define en *ventana*.

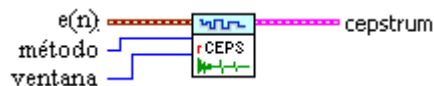


Figura 2.9. TSA Real Cepstrum.vi, utilizado en el cálculo del Cepstrum Real.

#### Detección de picos.

Los picos del cepstrum de la señal  $e(n)$  tienen una estrecha relación con el pitch de la señal de voz, por lo que su correcta detección resulta un proceso crítico en esta aplicación. Para el logro de la obtención de estos picos se utilizó un vi perteneciente también a la caja de herramientas Procesamiento Avanzado de Señales de LabVIEW 9.0 [33], WA Multiscale Peak Detection.vi mostrado en la Figura 2.10. Calcula la ubicación de los picos del cepstrum mediante análisis multiescala de wavelet [40], para esto utiliza un umbral de selección de los picos que fue prefijado en 0.99.

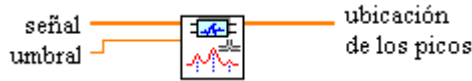


Figura 2.10. WA Multiscale Peak Detection.vi, utilizado en el cálculo de la ubicación de los picos del Cepstrum Real.

### 2.3.2. Método de detección de formantes mediante la extracción de picos espectrales utilizando LPC.

Mediante el estudio de la variación en las estructuras de los formantes se pueden determinar distintas patologías así como evaluar la evolución del paciente en un proceso de rehabilitación del habla.

Existe una serie de métodos para el cálculo de los formantes en las señales de voz, están basados en detección de picos espectrales, extracción de raíces y análisis mediante síntesis de las señales de voz [41]. Los métodos típicos de estimación de los formantes en el dominio de la frecuencia involucran la búsqueda de picos de representaciones de espectros usualmente con Transformada de Fourier de Tiempo Corto [24] o análisis mediante LP [13].

Los métodos de extracción de picos espectrales y sus variantes son usados frecuentemente para estos fines pues muestran resultados relativamente fiables, además de que poseen poco costo computacional [13]. Uno de los problemas de estos métodos es la obtención de espectros que sean adecuados para la correcta detección de los picos. El análisis mediante LP ofrece una alternativa para la obtención de un espectro útil en este caso, debido a su característica de lograr un suavizado de los picos espectrales, los picos mayores se obtiene bien definidos, mientras que otros, que podrían contribuir a resultados erróneos, son atenuados.

Para la obtención de los formantes en el software se utilizó un algoritmo basado en la detección de picos espectrales a partir de la obtención de los coeficientes LPC. En la Figura 2.11 se muestra un diagrama en bloques de este algoritmo, donde  $S(n)$  es la señal original,  $X(n)$  la tramas de la señal obtenida mediante un inventanado de  $S(n)$  por la ventana  $w(n)$ .

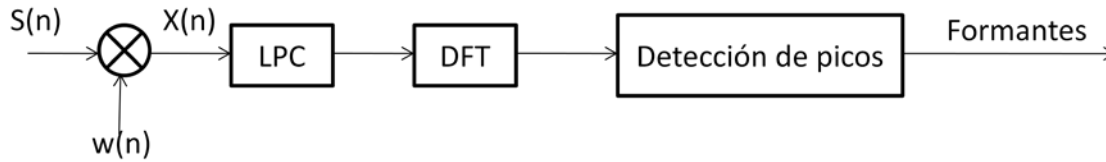


Figura 2.11. Diagrama en bloques de la obtención de formantes mediante el método de LPC,  $S(n)$  señal de voz,  $w(n)$  ventana Hanning de 40mseg,  $X(n)$  trama de señal enventanada. (Tomado de [36])

El proceso de división en tramas de la señal de voz se hace de manera análoga a la realizada en la obtención del pitch, usándose una ventana Hanning de 40mseg de duración. Luego de la obtención de los coeficientes LPC se obtiene un espectro de los mismos utilizando la Transformada Discreta de Fourier (DFT por sus siglas en inglés), a partir del cual se hace la detección de picos espectrales y se extraen los formantes.

Cálculo de LPC.

La obtención del LPC de las tramas de señal se realizó mediante la utilización de una herramienta del procesamiento gráfico de LabVIEW 9.0 [33], Figura 2.12, que realiza un modelado AR de la señal, esta herramienta se encuentra en el set de vi disponibles en la caja de herramienta de Procesamiento Avanzado de Señales de LabVIEW 9.0.

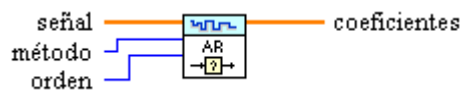


Figura 2.12. TSA AR Modeling.vi utilizado en el cálculo del LPC de la señal.

TSA AR Modeling.vi realiza un modelado Autorregresivo de acuerdo al método que se especifique en *método*. De esta forma al definir un método del tipo “Forward-Backward” o “Hacia delante-Hacia atrás” se calculan los coeficientes minimizando el error de la predicción. Como el análisis LPC establece un modelado de tipo todo-polos o AR y se basa en la optimización de los coeficientes minimizando el error de predicción se puede tomar que el proceso realizado por TSA AR Modeling es análogo al realizado en el análisis de LP y los coeficientes resultantes de esta operación pueden ser tratados como los coeficientes LPC de las tramas de la señal [32]. En *orden* se especifica el orden de predicción, en este caso se utilizó un valor de 20, este valor permite obtener un nivel más cercano de los picos espectrales a partir del LPC.

Cálculo de la DFT.

El cálculo del espectro modificado de la señal se hace utilizando una herramienta disponible en LabVIEW 9.0, mostrado en la Figura 2.13. Para señales de una dimensión FFT.vi calcula la DFT de  $X$ , utilizando el algoritmo de Transformada Rápida de Fourier (FFT por sus siglas en inglés), [24]. Se utilizó una longitud de la FFT igual a 40mseg.

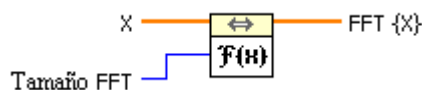


Figura 2.13. FFT.vi utilizado para el cálculo del espectro de los coeficientes LPC de la señal.  $X$  corresponde a los coeficientes LPC de  $x(n)$ .

### Detección de picos.

La detección de picos se hizo utilizando la misma herramienta que en el cálculo de la  $F_0$ , o sea, mediante análisis multiescala de wavelet [40], pero a diferencia del caso anterior, se especifica el ancho del pico a extraer en *ancho del pico*, Figura 2.14. Ambos valores, *umbral* y *ancho del pico*, pueden ser seleccionados a voluntad por el usuario, ya que fueron puestos en el panel frontal del software a modo de selectores para este propósito.

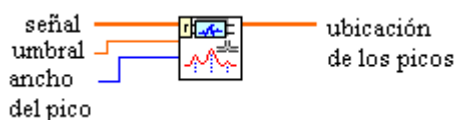


Figura 2.14. WA Multiscale Peak Detection.vi, utilizado en el cálculo de la ubicación de los picos espectrales para la extracción de los formantes.

Ambos algoritmos, el de extracción de  $F_0$  y el de detección de formantes se encuentran disponibles de forma gratuita en la página web Using LabVIEW for Voice Signal Analysis [36], de la National Instrument. Forman parte de un programa confeccionado en LabVIEW para análisis de señales de voz, Figura 2.15, que sirvió como punto de partida para la implementación del software de este trabajo. Aunque el software Voice Signal Analysis resulta útil para aplicaciones relacionadas con la extracción de características de las señales de voz, consideramos que su uso no resulta completamente extensible al caso de las consultas de logopedia y foniatría. Esto se debe, entre otras cosas, al hecho de que este programa consta de terminología técnica propia del procesamiento de voz que resultaría foránea al personal ajeno a esta rama de la ingeniería. Además de que está privado de otras



ventajas como la posibilidad de grabar una señal de voz, de reproducirla y de confeccionar un TV de un paciente, estrategias necesarias también para utilizar en esta aplicación.

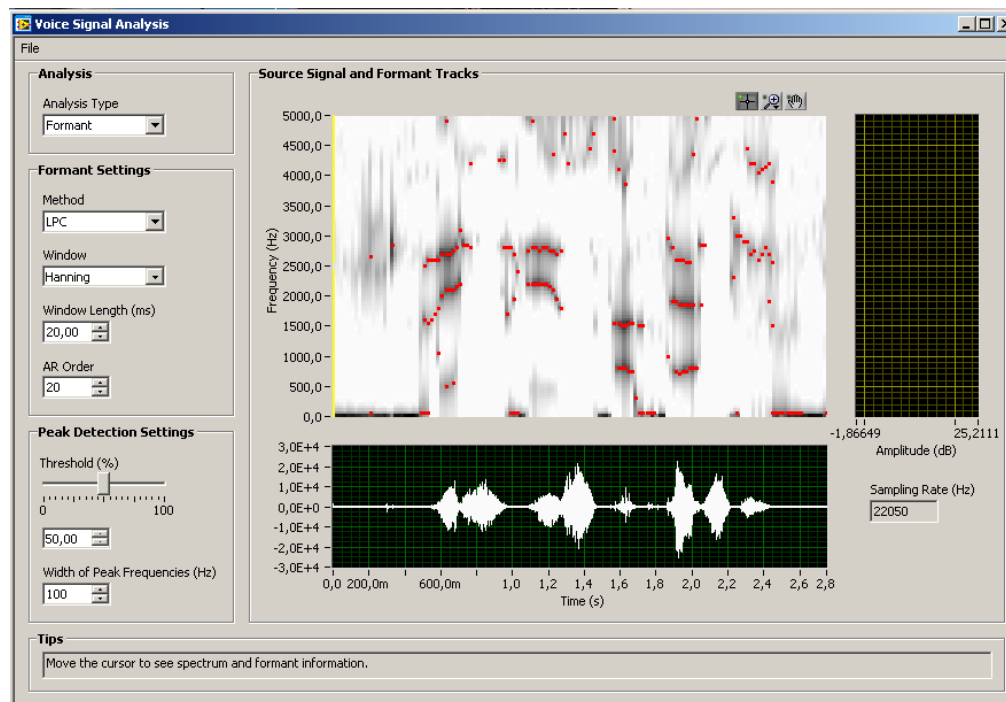


Figura 2.15. Panel frontal del programa Voice Signal Analysis obtenido en [36].

### 2.3.3. Construcción del Triángulo Vocálico y cálculo de su área.

Con el objetivo de obtener el TV se toman los valores de F1 y F2 seleccionados por el operador por medio de un cursor situado en el gráfico de formantes del programa, para cada vocal sostenida (/a/, /i/, /u/) del paciente. Por medio de un gráfico estos valores son ubicados uno a uno en un eje de coordenadas (F1, F2), hasta completar el TV. El software brinda la posibilidad de guardar el triángulo formado o buscar otro guardado anteriormente, esto resulta en una ventaja tanto para el médico como para los pacientes, ya que hace posible de manera real y cualitativa la comparación de las características vocales de un individuo en un proceso de rehabilitación o aprendizaje del habla, como en el caso de los pacientes con implantes cocleares.

El cálculo del área del TV es otra característica significativa que puede ser empleada para diferenciar entre voces normales y patológicas, además de que permite dar prueba del desempeño de un paciente en un proceso de rehabilitación. Habiendo construido el

triángulo con anterioridad no resulta complicada la obtención de su área, mediante la utilización de la fórmula de Herón, Ecuación 11 y el Teorema de Pitágoras, Ecuación 13, podemos calcularla de la siguiente forma:

$$A = \sqrt{p(p-a)(p-b)(p-c)} , \quad 11$$

Donde  $p$  es el semiperímetro del triángulo y  $a$ ,  $b$  y  $c$  los lados del triángulo. El semiperímetro se obtiene según la Ecuación 12.

$$p = \frac{(a+b+c)}{2} , \quad 12$$

El Teorema de Pitágoras permite el cálculo de la hipotenusa de un triángulo rectángulo a partir de conocer los valores de sus catetos correspondientes. La hipotenusa,  $m$ , se obtiene como enuncia la Ecuación 13, donde  $n$  y  $l$  son los catetos del triángulo.

$$m^2 = n^2 + l^2 , \quad 13$$

Si hacemos extensivo el uso del Teorema de Pitágoras a este caso podemos calcular las longitudes de los lados del TV, Figura 2.16, a partir de los valores de  $F1$  y  $F2$  pertenecientes a las vocales ubicadas en los extremos de cada lado. Los lados del TV corresponden a la hipotenusa de un triángulo rectángulo cuyos catetos están formados por las diferencias de los formantes de las vocales en los puntos extremos de cada lado. De esta forma, la longitud del lado  $a$ , se obtendría a partir de la Ecuación 14.

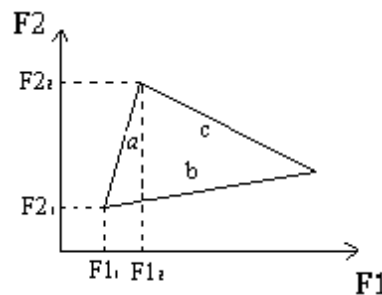


Figura 2.16. Ejemplo de TV. Descomposición en coordenadas de las vocales de los extremos del lado  $a$ .

$$a = \sqrt{(F2_2 - F2_1)^2 + (F1_2 - F1_1)^2}, \quad 14$$

Para obtener la longitud de b y c se procese de la misma forma, solo que cambian los valores de F1 y F2 por los valores correspondientes a cada caso.

#### 2.4. Adquisición y reproducción de las señales de voz.

El proceso de grabación se hace utilizando un micrófono y mediante la tarjeta de sonido de la computadora, con una frecuencia de muestreo fija de 22050Hz, esta frecuencia garantiza una mejor calidad de sonido, además, como las consultas de logopedia y foniatría hacia las que va dirigida la utilización de este software, constan de un ambiente de ruido regulado, no existe un riesgo alto de contaminación de la señal de voz por ruido ambiental. El archivo de la señal de voz es guardado utilizando la extensión “.wav” en una dirección anteriormente especificada por el usuario.

La reproducción se hace desde una señal que se haya guardado anteriormente bajo la extensión “.wav” y utilizando también la tarjeta de sonido que posea la computadora, además de bocinas o audífonos que posibiliten escuchar la grabación. Al cargar una señal previamente grabada, esta puede ser reproducida y realizar el proceso de extracción de variables características que ofrece el software.

#### 2.5. Conclusiones parciales.

Los algoritmos anteriormente explicados fueron utilizados en la conformación de un software nombrado FORMANTE, implementado en LabVIEW 9.0. Estos algoritmos son relativamente sencillos en su construcción, además permiten la utilización de herramientas novedosas que brinda esta versión de LabVIEW; lo que hace del software una herramienta rápida y eficiente en la caracterización de señales de voz.

## CAPÍTULO 3: RESULTADOS Y DISCUSIÓN.

### 3.1. Software FORMANTE.

A partir de los algoritmos explicados anteriormente, utilizando el programa LabVIEW 9.0 [33], se implementó un software cuya interfaz gráfica se muestra en la Figura 3.1; nos brinda facilidades como grabar una señal de voz y reproducir una que se haya guardado anteriormente. Además permite obtener los formantes de una señal y su correspondiente oscilograma, la frecuencia fundamental y construir un TV a partir de las vocales sostenidas /a/, /i/, /u/, guardarlo y visualizar otro que haya sido guardado anteriormente por el programa, además de calcular el área del mismo. Se ha incluido en la instalación una ayuda detallada de la utilización del mismo, para facilitar su comprensión y utilización.

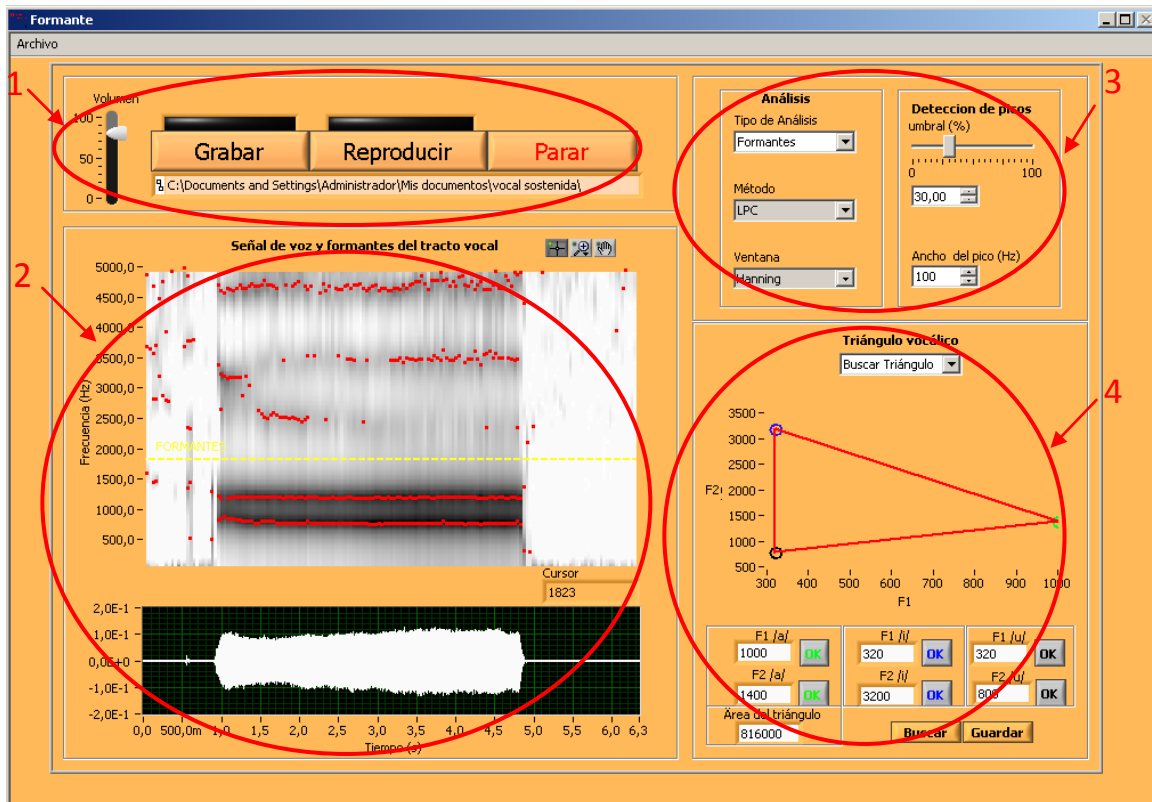


Figura 3.1. Interfaz gráfica del software “Formante”, implementado en este trabajo. 1. Botones de control y barra de direccionamiento. 2. Gráficas. 3. Controles para extracción de F0 y formantes. 4. Sección de Triángulo Vocálico.

El software cuenta con un archivo instalador y un ejecutable que posibilita la utilización del mismo en sistemas operativos como Windows, desde 2000 en adelante, sin que sea

necesario tener instalado el programa LabVIEW 9.0 en la máquina. Esto resulta una ventaja considerable debido al hecho de que las computadoras de los usuarios potenciales del software FORMANTE no poseen características de alto procesamiento y rendimiento como para utilizar el software LabVIEW 9.0 en su totalidad.

Para una mejor comprensión de la utilización del software se dividió la interfaz gráfica en cuatro secciones señalizadas en rojo en la Figura 3.1:

1. Botones de control y barra de direccionamiento.
2. Gráficas.
3. Controles para extracción de F0 y formantes.
4. Sección de Triángulo Vocálico.

En la sección 1 se encuentran los botones relacionados con la adquisición de la señal y su reproducción, como son: GRABAR, REPRODUCIR, PARAR, el control de volumen y la barra de direccionamiento donde se visualiza la dirección del archivo de voz sobre el que se está trabajando.

La sección 2 contiene las gráficas de forma de onda, así como las correspondientes a frecuencia fundamental y formantes, estas últimas son mostradas de acuerdo al tipo de análisis que se efectúa, que puede ser seleccionado con el control “*Tipo de Análisis*” en la sección 3.

La sección 3 contiene dos indicadores que muestran el método de obtención de los formantes (LPC) y el tipo de ventana que se utilizó para la división en tramas de la señal (Hanning). Además cuenta con el control “*Tipo de Análisis*” que permite seleccionar entre un análisis de formantes y uno de frecuencia fundamental. Para el caso de análisis de los formantes de la señal, esta sección también dispone de un selector del umbral utilizado para la detección de los picos espectrales, que aparece en un valor de 30% por defecto, también permite escoger el ancho de banda del pico, en Hz, prefijado por defecto en un valor de 100Hz.

La sección 4 corresponde a la confección del triángulo vocálico. Consta de una gráfica donde se visualiza el TV, así como un indicador donde se muestra el área del mismo. La utilización de esta sección será explicada posteriormente con más detalle.

Además este programa consta de una barra de menú y una barra de título. La barra de menú contiene el selector “*Archivo*”, Figura 3.2. Este permite cargar una señal de voz, mostrar el cuadro de ayuda contextual y salir del programa seleccionando “*Exit*”.

Para proceder a cargar una señal se selecciona “*Cargar señal*” en el menú “*Archivo*”, a continuación se mostrará un cuadro de diálogo que permite buscar la dirección de la señal que se desea analizar y seleccionarla. Para realizar cualquier tipo de análisis de los brindados por el software FORMANTE es necesario tener cargada la señal a analizar con anterioridad. La señal seleccionada se utilizará en el programa hasta tanto no sea seleccionada otra.

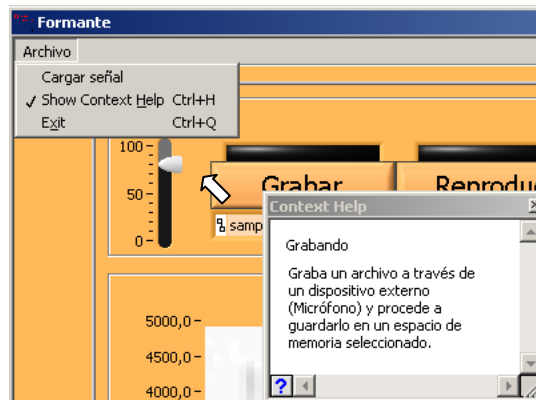


Figura 3.2. Representación del menú “*Archivo*”. Selección de “*Show Context Help*” y muestra de la ayuda contextual del programa.

El proceso de grabación se hace de acuerdo al diagrama de flujo mostrado en la Figura 3.3 a), mediante la utilización de un micrófono y la tarjeta de sonido de la computadora. Para detener el proceso se debe pulsar el botón PARAR de lo contrario este seguirá indefinidamente.

En el caso de la reproducción de la señal, hay que proceder primero a cargar la misma al programa, para luego ser reproducida y analizada de acuerdo a las facilidades que brinda el software, en la Figura 3.3 b) se muestra el diagrama de flujo para el proceso de reproducción. El control de volumen ubicado en la sección 1 permite aumentar o disminuir el nivel de intensidad del sonido en la reproducción del mismo.

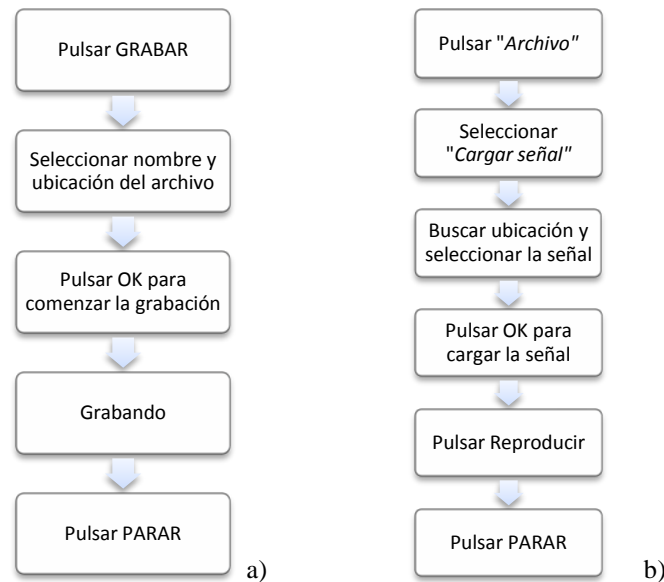


Figura 3.3. Diagrama de flujo del proceso de: a) Grabar la señal y b) Reproducir.

La construcción del TV resulta un poco más compleja para el usuario del software, ya que la selección de los formantes F1 y F2 de las vocales se hace de manera manual para cada una. Es necesario tener almacenada con anterioridad las vocales sostenidas /a/, /i/, /u/ de un paciente para la construcción del triángulo. El TV se va modificando a medida que se le vaya incorporando la información acerca de los valores de formantes de las vocales, el proceso es mostrado en la Figura 3.4. Además el software brinda también la posibilidad de calcular el área del mismo, mostrando el valor en el indicador “Área del triángulo”, ver sección 4 de la Figura 3.1. Las vocales están diferenciadas por colores, el verde pertenece a la /a/, el azul a la /i/, mientras que la /u/ está señalada con color negro, así resulta mucho más fácil encontrar su ubicación en el triángulo vocálico.

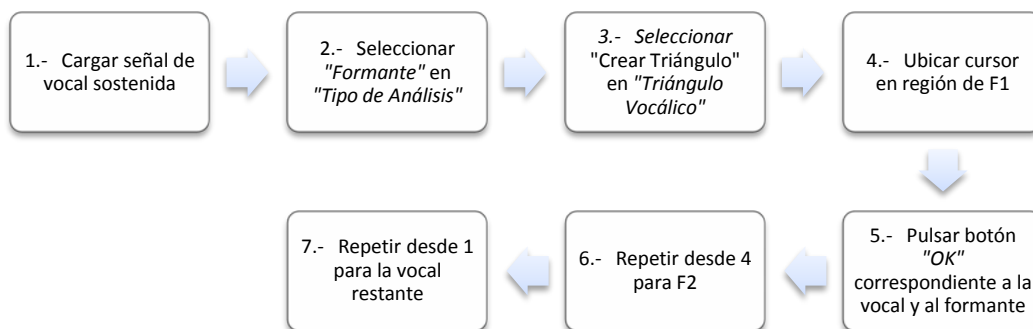



Figura 3.4. Diagrama de flujo para el proceso de construcción del Triángulo Vocálico.

Luego de haber confeccionado un triángulo se puede guardar el mismo al hacer clic en el botón GUARDAR de la sección 4; especificando la dirección y el nombre del archivo. El triángulo guardado puede ser visualizado nuevamente en la interfaz del software si se selecciona “*Buscar triángulo*” en el menú “*Triángulo Vocálico*” y posteriormente se pulsa el botón BUSCAR de la sección 4 para encontrar el archivo del triángulo que se desea mostrar. Esta utilidad del software permite hacer comparaciones del estado del paciente a lo largo de un proceso de rehabilitación o de diagnóstico.

Para obtener el gráfico de formantes o frecuencia fundamental se selecciona la opción correspondiente en el selector “*Tipo de Análisis*”. El control “*umbral*” situado en la región “*Detección de picos*” de la sección 3 permite modificar el umbral en el proceso de extracción de picos espectrales para la obtención de los formantes. De igual forma se puede modificar el ancho de banda de los picos a extraer, modificando el control, “*Ancho del pido (Hz)*” ubicado en la misma sección. Esto permite acomodar el gráfico de los formantes según le sea más satisfactorio al usuario. La extracción de los formantes se hizo utilizando el método explicado en el Capítulo 2, basado en LPC, mientras que la extracción de la frecuencia fundamental se hizo utilizando el método Cepstrum explicado también en el Capítulo 2.

En el caso de que se necesite información sobre la utilización del programa este consta de un cuadro de ayuda contextual que brinda una descripción detallada de los componentes del panel frontal. Para tener acceso a esta utilidad se selecciona “*Show Context Help*” en el menú “*Archivo*”. Así se mostrará una ayuda contextual de cada elemento indicado por el cursor. Para ampliar la información que ofrece el cuadro de ayuda contextual se pulsa el icono  que nos ofrece una ayuda general del programa. Como se muestra en la Figura 3.2.

El gráfico de los formantes puede ser usado de manera efectiva en la detección de distintas patologías, así como en la evaluación del desempeño del paciente en un proceso de rehabilitación. Las regiones correspondientes a los formantes son destacadas en color rojo por el proceso de detección de los mismos. En la Figura 3.5 se muestran dos gráficos correspondientes a una vocal /a/ sostenida de una voz masculina normal, a la izquierda y patológica, a la derecha. En este caso la patología presenta características de ronquera, que provocan que haya una contaminación de la señal con componentes de ruido, lo que



enmascara notablemente la aparición de los formantes, esto puede comprobarse en dicha figura.

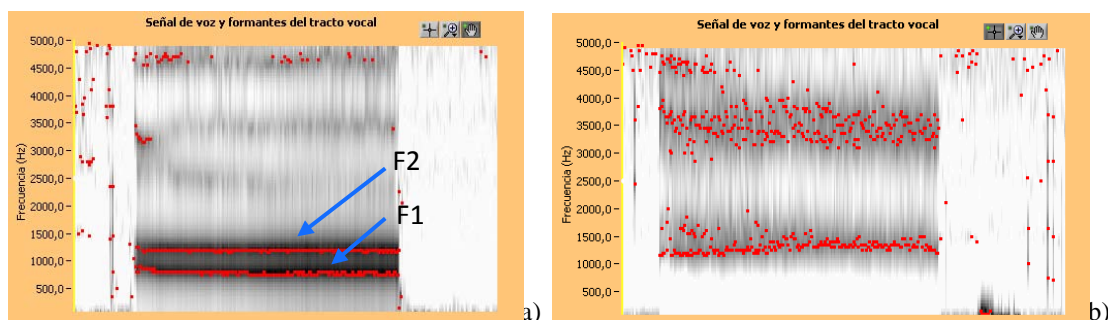


Figura 3.5, Gráfico de formantes de vocal /a/ sostenida de voz masculina, a) normal y b) patológico.

En la Figura 3.5 a) están señalados los formantes F1 y F2 de la vocal /a/, ambos se encuentran estables y bien definidos, además sus valores están dentro de los clasificados como normales, ver Figura 1.5, por lo que puede demostrarse que este gráfico corresponde a una voz normal. Sin embargo en la Figura 3.5 b) no es posible definir cuáles son las regiones correspondientes a los formantes, el espectro presenta notables afectaciones por ruido característico de este tipo de patología.

De igual forma se puede notar la diferencia entre sonidos utilizando los formantes, en la Figura 3.6 se muestra el gráfico de formantes de una voz femenina normal pronunciando tres vocales (/a/, /i/, /u/), sostenidas consecutivamente. En el gráfico puede verse como los formantes detectados se corresponden con sus valores normales esperados, ver Figura 1.5.

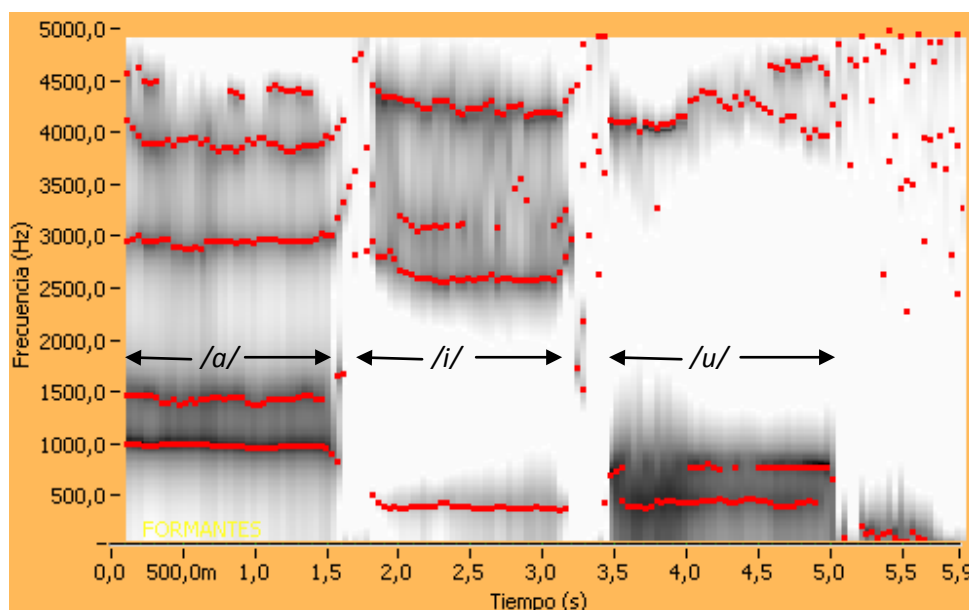


Figura 3.6. Gráfico de formantes de una voz femenina pronunciando las vocales /a/, /i/, /u/ consecutivamente.

Si realizamos la construcción del triángulo vocálico correspondiente al caso de la Figura 3.6 según se ha explicado con anterioridad, se obtiene un triángulo vocálico normal mostrado en la Figura 3.7. El TV puede verse diferenciado en presencia de una patología como en el caso de que exista un problema de articulación que provoque una diferenciación pobre entre los sonidos vocálicos. Esto puede ocasionar una centralización del triángulo vocálico, disminuyéndose el valor de su área.

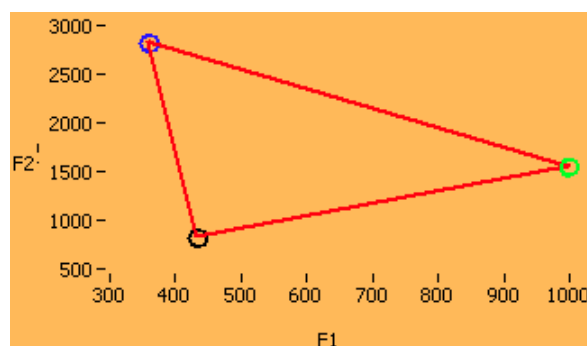


Figura 3.7. Triángulo Vocálico normal construido en el software a partir del gráfico de formantes de la figura 3.6. El color verde corresponde a la vocal /a/, el color azul a la /i/ y el negro a la vocal /u/.

La frecuencia fundamental también puede ser usada para diferenciar entre una voz normal y una patológica. Esta característica de la voz varía según el estado de salud del paciente, su edad, sexo, etc. En la tabla 1 se muestran valores de frecuencia fundamental obtenidos a partir de gráficos del software correspondientes a varias voces normales, masculinas, femeninas, adultas e infantiles, pronunciando la vocal /a/ sostenida. Los valores se obtuvieron de forma aproximada mediante inspección visual de los gráficos, se muestra el valor medio y el rango de variación de F0 para cada voz.

Tabla 1. Valores aproximados del promedio y rango de variación de F0 de varias voces normales pronunciando la vocal /a/ sostenida.

Señales de voz	Sexo	Edad	Valor medio F0 (Hz)	Rango de variación de F0 (Hz)
Voz-1	Femenino	48	203	202-212
Voz-2	Masculino	52	179	178-182
Voz-3	Femenino	8	263	259-269
Voz-4	Femenino	23	240	237-248
Voz-5	Masculino	7	242	234-248

Se muestra en la Tabla 1 como los valores obtenidos mediante el gráfico permiten diferenciar entre distintas voces. Además la detección de F0 en voces normales como estas, presentan un estrecho rango de variación individual, sin importar el tipo de voz que se analice. Esto demuestra la efectividad del algoritmo detector de F0.

En presencia de una patología, la frecuencia fundamental puede verse afectada de diferentes maneras. En la Figura 3.8 se muestran dos gráficos de frecuencias fundamentales pertenecientes a dos voces masculinas: sana y patológica. Los gráficos pertenecen a un análisis de F0 de las voces utilizadas para la confección de la Figura 3.5. En la Figura 3.8 se observan las diferencias entre las F0 de ambos casos, en la voz normal, a), tiene un valor prácticamente constante a lo largo del tiempo mientras que en el caso patológico, b), resulta imposible hacer una detección correcta del valor de F0 del paciente, debido a la inestabilidad que presenta esta característica, producto a la contaminación por ruido propia de la voz ronca del individuo.

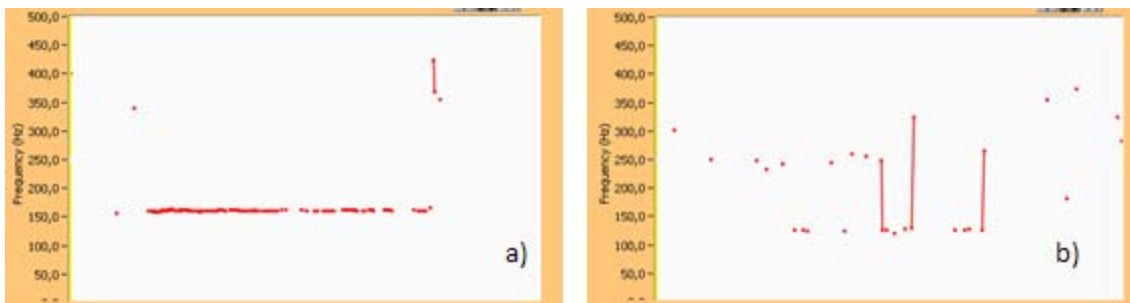


Figura 3.8. Gráfico de frecuencias fundamentales, vocal /a/ sostenida, masculina, a) normal y b) patológica.

### 3.2. Conclusiones parciales del capítulo.

En este capítulo se comprueba que el software cumple con los requerimientos y especificaciones esperadas. La interfaz posee una serie de controles e indicadores de fácil manejo para el usuario y se explica detalladamente una metodología para la utilización de la aplicación. Además se muestran algunos ejemplos del comportamiento del mismo ante distintas señales de voz.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones:

- Se seleccionó el algoritmo adecuado para la determinación del período fundamental.
- Se seleccionó el algoritmo adecuado para la determinación de formantes.
- Se realizó la puesta a punto del software.
- Se diseñó y desarrolló un software para el registro de características de la voz utilizable en consultas de logopedia y foniatría.

### Recomendaciones:

- Promover la aplicación del software propuesto en las consultas de logopedia y foniatría de la región con el objetivo de obtener una primera realimentación de los resultados y deficiencias del mismo, que contribuyan a su posterior desarrollo.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] M. Mohan and J. Schroeter, "Speech Production Models and Their Digital Implementations," in *Digital Signal Processing Handbook*, V. K. Madisetti and D. B. Williams, Eds., 1999, ch. 44.
- [2] César San Martín and Roberto Carrillo, "Implementación de un reconocedor de palabras aisladas dependiente del locutor.," *Revista Facultad de Ingeniería, U.T.A (Chile)*, vol. 12, no. 1, pp. 9-14, 2004.
- [3] Arthur C & Hall Guyton, "Respiración.," in *Textbook of medical Physiology.*, ch. VII, pp. 537-538.
- [4] Un Chong and Yang Shih-Chien, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 6, pp. 565 - 572, Diciembre 1977.
- [5] Hema A. Murthy and K. V. Madhu Murthy, "Formant Extraction from Fourier Transform Phase," *IEEE Trans. on Biomedical Engineering*, pp. 484-488, 1989.
- [6] P. Martín, "Comparison of pitch detection by cepstrum and spectral comb analysis," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, pp. 180-183, Mayo 1982.
- [7] María E. Hernández-Díaz, "Algoritmos para la Extracción del Período Fundamental de la Voz: Desarrollo y Evaluación," Universidad Central Martha Abreu de Las Villas, Santa Clara, Tesis doctoral 1996.
- [8] H. & Hernández-Díaz, M. , Kairuz, "Análisis del Triángulo Vocálico en pacientes con implante coclear," IV Encuentro Iberoamericano de Trastornos del Lenguaje, Habla y Voz y Congreso Cubano, Habana, Cuba., 2008.
- [9] Eduardo Castillo, "Algoritmos y Programas para el estudio de voces patológicas disártricas.," Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI), Universidad Central de Las Villas (UCLV), Santa Clara, Cuba, 1996.
- [10] Héctor A. Kairuz Hernández – Díaz, "Detección y marca de los cambios espectrales notorios en ] voces patológicas," Centro de Estudios de Electrónica y Tecnologías de la Información (CEETI), Universidad Central de Las Villas (UCLV), Tesis de grado 2009.
- [11] Matías Zañartú, "Aplicaciones del análisis acústico en los estudios de la voz humana," ] Seminario Internacional de Acústica, Santiago de Chile, 2003.

- [12 M. J. Llau Arcusa and J. González, "Medida de la inteligibilidad en el habla disártrica," *Revista de Logopedia, Foniatría y Audiología*, vol. 24, no. 1, pp. 33-43, 2004.
- [13 Susanna Varho, *New Linear Predictive Methods for Digital Speech Processing*. Helsinki, 2001.
- [14 Alex Acero, "Source-Filter Models for Time-Scale Pitch-Scale Modification of Speech," Microsoft Research, Wasington,.
- [15 F. A., Castellanos G. Sepúlveda, "Estimación de la Frecuencia Fundamental de Señales de Voz Usando Transformada Wavelet," 2004.
- [16 WJ Hess, "Pitch and Voice Determination of Speech with and Extesion Toward Musical Signal," in *Springer Handbook of Speech Processing.*, 2008, ch. 10, pp. 181-208.
- [17 J.A. Facal, M.L. y Salvatori, J.M. Gurlekian, "Preparación de un Informe Clínico para tres patologías de la voz.," in *Topicos em Fonoaudiología. CEFAC*, Zorzi, Gomes Marchesan, Ed.: LOVISE, pp. Pp. 539-554.
- [18 Jose Antonio Díaz and Christine Sapienza, "Algoritmo robusto para la detección de la frecuencia fundamental de la voz basado en el espectrograma," *Revista de Ingeniería UC*, vol. 10, no. 3, pp. 7-16, 2003.
- [19 M., Ramón, J. L. Rosique, "Análisis Discriminante Aplicado a los Formante de las Vocales Castellanas en la Fonación con Prótesis y Erigmo fonía tras Laringuectomía total," España, 2003.
- [20 R. D., Weismer, G., Kent, J. F., Vorperian, H. K. Kent, "Acoustic Studies of Dysarthic Speech: Methods, Progress, and Potentia," University of Wisconsin Madison, 1999.
- [21 B.J., & Weisiger, B.E. Zski, "Identification of dysarthia types based on perceptual analysis," *Journal of Communication Disorders*, vol. 20, pp. 367-378, 1987.
- [22 R. J. Baken, *Clinical Measurement of Speech and Voice*. San Diego, 1999, ch. 9.
- [23 R. W. Schafer, "Homomorphic System and Cepstrum Analysis of Speech," in *Springer Handbook of Speech Processing*. Canadá, 2008, ch. 9, pp. 161-180.
- [24 W. Keneth Jenkins, "Fourier Series, Fourier Transforms, and DFT," in *Digital Signal Processing*, V. K. Madisetti and D. B. Williams, Eds., 1999, ch. 1.
- [25 A. MICKAEL NOLL, "Short-Time Spectrum and "Cepstrum" Techniques for Vocal Pitch

- ] Detection," *The Journal of the Acoustical Society of America*, vol. 36, no. 2, pp. 296-302, Febrero 1964.
- [26 Alan V. Oppenheim and Ronald W. Schafer, "From Frequency to Quefrency: A History of the  
] Cepstrum," *IEEE Signal Processing Magazine*, pp. 95-106, Septiembre 2004.
- [27 Michael A. Noll, "Cepstrum Pitch Determination," *The Journal of the Acoustical Society of  
] America*, vol. 41, no. 2, 1967.
- [28 H. Laforcada and D. Milone, "Diseño y desarrollo de un Software para el análisis y  
] procesamiento de señales de voz.," in *Memorias del XIV Congreso Argentino de Bioingeniería*, 2003, p. 66.
- [29 H. Kobayashi and T. Shimamura, "A modified cepstrum method for pitch extraction," in *The  
] 1998 IEEE Asia-Pacific Conference on Circuits and Systems. IEEE APCCAS*, Chiangmai , Thailand, Noviembre 1998, pp. 299-302.
- [30 B. S. Atal, "The history of linear prediction," *IEEE Signal Proc.* , no. 23, pp. 154-161, 2006.  
]
- [31 J. Benesty, J. Chen, and Y. Huang, "Linear Prediction," in *Springer Handbook of Speech  
] Processing*. Canadá, 2008, ch. 7, pp. 121-133.
- [32 Jhon G. Proakis and Dimitris G. Manolakis,,: Prentice Hall, 1998, ch. 11, pp. 861-900.  
]
- [33 National Instrument, Labview V9.0, 2009.  
]
- [34 David Gerhard, "Pitch Extraction and Fundamental Frequency: History and Current  
] Techniques," Department of Computer Science University of Regina, Regina, Canada, 2003.
- [35 Gastón Schlotthauery, María E. Torres, and Hugo R. Rufiner, "Determinación de la frecuencia  
] fundamental de la voz basada en descomposición modal empírica por conjuntos y entropías," *XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC 2009)*, Septiembre 2009.
- [36 (2011, Febrero) Using LabVIEW for Voice Signal Analysis. [Online]. HYPERLINK  
] "http://zone.ni.com/devzone/cda/epd/p/id/5977" <http://zone.ni.com/devzone/cda/epd/p/id/5977>
- [37 Hiroshi Muta and Thomas Baer, "A Pitch-synchronous Analysis of Hoarseness in Running  
] Speech," Haskins Laboratories, 1988.

- [38 Sassan Amhadi, "New Techniques for Sinusoidal Coding of Speech al 2400 BPS," *IEEE*, 2006.  
]
- [39 Hannu Pulakka, "Analysis of Human Voice Production Using Inverse Filtering, High-Speed  
] Imaging, and Electroglottography," HELSINKI UNIVERSITY OF TECHNOLOGY, Tesis de maestría  
2005.
- [40 M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*. Massachusetts: A K  
] Peters, 1994.
- [41 Chamwoo Kim, "A Robust Formant Extraction Algorithm Combining Spectral Peak Picking and  
] Root Polishing," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1-16, 2006.
- [42 E. and Gath, I. Yair, "High resolution pole - zero analysis of Parkinsonian speech," *IEEE Trans.*  
] *on Biomedical Engineering*, vol. 38, no. 2, 1991.
- [43 J. L. and Del Riesgo, C. Flanagan, "Speech Processing: A perspectiveon the science and its  
] applications," *AT & T Technical Journal*, pp. 2-13, Sept./ Oct 1990.
- [44 Eduardo Castillo and Carlos Ferrer, "Caracterización de rupturas del período en voces  
] patológicas," *Memorias II Congreso Latinoamericano de Ingeniería Biomédica*, Mayo 2001.
- [45 V., Jamieson, D. G., Parsa, "A Comparison of High Precision FO Extraction Algorithms for  
] Sustained Vowels," University of Western Ontario. , Canada, 1999.
- [46 María G. Sánchez, "Desórdenes motores del habla y PROMPT".  
]
- [47 J. Mohan Sondhi, M., Yiteng H. Benesty, *Springer Handbook of Speech Processing*. Canadá,  
] 2008.
- [48 V. K., Williams, D. B. Madisetti, *Digital Signal Processing Handbook.*, 1999.  
]
- [49 MathWorks Inc. (2009) Matlab. V7.9.  
]