

1 RESUMEN

El siguiente trabajo de tesis fue desarrollado durante el período comprendido entre los meses de Septiembre del 2003 y Mayo del 2004. Al principio de dicho período fue necesario reunir antecedentes teóricos adecuados para adquirir un lenguaje, o los conocimientos necesarios para desarrollar nuestro trabajo. Vale la pena destacar que gracias a la ayuda del equipo multidisciplinario formado por docentes de nuestra casa de estudios y la Universidad Mayor fue posible, desde nuestro punto de vista, cumplir con los objetivos propuestos.

La existencia de trabajos de investigación en Fonética realizados en otros países y el hecho de que el análisis acústico ha tomado mayor importancia en los últimos años tanto para el área de la Fonoaudiología como para el área de la Acústica como tal en Chile, son antecedentes necesarios que invitan a innovar y desarrollar un proyecto donde la participación de profesionales de dichas áreas es indispensable.

Este trabajo presenta el desarrollo de una Base de Datos Fonética para voces del territorio chileno, que sea efectivo, confiable y de bajo costo, y que a la larga permita obtener metodologías de trabajo en las diversas áreas de la Fonética, específicamente en el área de la Fonética Acústica.

La construcción de la Base de Datos implicó criterios ya existentes de diseños y bajo parámetros preestablecidos, es decir, que nos permitan tener un acceso a los archivos de audio de acuerdo a las clasificaciones realizadas.

Se presentan además las metodologías de análisis en las principales áreas de aplicación, como son la determinación del Tono Medio Hablado, cálculo de Formantes, parámetros de Voz Normal y Voz Patológica, y el análisis de la Prosodia.

2. INTRODUCCIÓN

2.1 Formulación del problema

El análisis acústico de la voz ha alcanzado un importante desarrollo en los últimos tiempos gracias, entre otras razones, al progreso y difusión experimentados por los medios informáticos que lo hacen posible. Entre sus ventajas, destaca el ser un método no invasivo de evaluación de la voz y el ofrecer la oportunidad de objetivizar la evaluación de ciertos parámetros acústicos.

La Universidad Pérez Rosales en conjunto con la Escuela de Fonoaudiología de la Universidad Mayor han iniciado un proyecto de cooperación académica con el objetivo de desarrollar investigación científica en las áreas de Fonética – Acústica y de patologías del habla.

De ésta manera surge la necesidad de contar con un registro sonoro que impulse la elaboración de tesis de pre-grado, la generación de equipos multidisciplinarios de trabajo entre ambas universidades y proyectos con otras casas de estudios.

La Base de Datos Fonética de Voces Chilenas pretende convertirse entonces en el primer registro de hablantes a nivel nacional, creada con el propósito de potenciar el desarrollo de la acústica en el área de la voz humana.

Finalmente, con la creación de ésta Base de Datos de información de hablantes chilenos, se puede dar pie ha numerosos estudios fonéticos acústicos enfocados hacia nuestro país, ya sea como el estudio o evaluación de la articulación vocálica o de la prosodia en el medio chileno, también se establecerá un precedente para nuevos proyectos en ésta área de las Universidades implicadas (U. Pérez Rosales, U. Mayor).

2.2 Objetivos Generales

- ✓ Construir una Base de Datos Fonética de voces chilenas mediante una interfaz telefónica para ser utilizada en futuros estudios en el área Lingüística, de Voz Normal y Patológicas.
- ✓ Demostrar las utilidades de la Base de Datos Fonética en cada una de las áreas.

2.2.1 Objetivos Específicos

- ✓ Implementar un sistema electroacústico para realizar la Base de Datos.
- ✓ Determinar el número de personas que serán grabadas en la Base de Datos.
- ✓ Estructurar el contenido del instructivo que se les enviará a las personas anteriormente seleccionadas, considerando números y fonemas de la lengua española.
- ✓ Diseñar una metodología que permita realizar las grabaciones.
- ✓ Obtener un registro sonoro de las muestras.
- ✓ Realizar una edición de las muestras obtenidas.
- ✓ Clasificar las muestras de acuerdo al tipo de estudio al que serán orientados (tono medio hablado, fonemas y coarticulación, voz normal, voz patológica y prosodia).

2.3 Hipótesis de Trabajo

- ✓ La construcción de la Base de Datos fonética de voces chilenas entregará una representación de las características acústicas de los hablantes chilenos.

- ✓ Existen diferencias fonoaudiológicas (lingüísticas, fonéticas, voz normal) en las voces chilenas en función de criterios definidos (sociales, regionales, edad) que pueden ser detectados y analizados mediante muestras de audio captadas con una interfaz telefónica.

3 ANTECEDENTES

3.1 Introducción

El sistema de producción de habla no forma parte estricta del sistema sensorial humano, pero su importancia es indudable. La comunicación humana surgió en el momento en el que nuestros antepasados, en su lucha por la supervivencia y en respuesta a sus instintos se vieron obligados a transmitir a quienes les rodeaban, sus impresiones, sentimientos, emociones. Para ello se valieron de la mímica, de los gritos y las interjecciones, lo que constituyó un lenguaje biológico, y posteriormente surgió el lenguaje hablado y las manifestaciones pictóricas.

De tal modo la **palabra** es importante, que el signo diferencial entre el animal que siente y se mueve, y el hombre, que también goza de movimiento y de sensibilidad, radica en la **palabra**. La creación inanimada suena; el animal, jugando con el instinto, grita; sólo el hombre, articulando la voz, pronuncia y emite la palabra.

La comunicación oral, en el más amplio sentido, es la expresión de nuestros pensamientos por medio de la palabra hablada y con fines comunicativos, y tiene algunas ventajas prácticas con respecto a la comunicación escrita:

- ✓ **Por su facilidad:** Es el mecanismo natural de comunicación humana.
- ✓ **Por el aprendizaje:** Es el mecanismo más precoz de comunicación.
- ✓ **Por la sencillez:** Incluso aquellos que no saben leer ni escribir, saben hablar su propia lengua.
- ✓ **Por la capacidad expresiva:** Con la ayuda de la entonación (y de elementos gestuales), somos capaces de transmitir la carga emotiva deseada.

Más aún, si comparamos lo que hemos aprendido por experiencia directa y lo que hemos aprendido de nuestras relaciones y comunicación con los demás, podemos darnos cuenta de lo limitados que estaríamos sin poder comunicarnos.[18]

3.2 Base de Datos

Toda la información de señales de voz necesarias para la implementación de un determinado sistema o servicio se le llama **Bases de Datos**.

Para el caso de bases de datos donde se consideren locutores, éstas deben recoger la mayor variabilidad de situaciones posibles que se puedan dar durante la vida útil del propio sistema. Sin embargo, esto va a ser imposible de realizar, ya sea por el costo humano y económico que implica y por la imprevisibilidad de multitud de situaciones que se pueden dar durante el funcionamiento del sistema.

Si se analizan el contenido de las bases de datos de locutores que se utilizan normalmente, se puede apreciar que existe una gran disparidad de contenidos y finalidades. Sin embargo hay algunos factores que deben ser considerados en toda base de datos de locutores y que normalmente se encuentran presentes.

Una de las características importantes es la grabación multisesión, es decir, la realización de distintas tomas de la voz de los locutores, separadas en el tiempo, de forma tal que sea posible abordar el problema de la variabilidad temporal. Otra característica deseada es el equilibrio fonético, de manera que los distintos sonidos recogidos se correspondan lo mejor posible con las distribuciones habituales para cada lengua, por lo que la elaboración del contenido lingüístico de una base de datos debe ser extremadamente cuidadosa.

Otros factores, pero más generales, como la adecuación en las condiciones del canal y ruido entre las fases de entrenamiento y funcionamiento, el balance entre las edades y sexos de usuarios, condiciones del habla (leída, espontánea,

etc...), el agrupamiento y selección de los locutores, son también tomados en cuenta con mayor o menor importancia en función del objetivo que se pretenda conseguir con la respectiva base de datos que se este grabando.

Existen un gran número de bases de datos en el mundo relacionadas con éste tema. En Europa se realizan grandes esfuerzos para tener un registro de una gran diversidad de lenguas que se dan en el continente europeo.

Específicamente en España, existen pocas bases de datos que estén disponibles públicamente, y menos bases de datos que tengan relación con locutores. Entre las últimas realizadas, se tiene a la **TelVoice**, que recoge habla telefónica de decenas de persona hasta en diez sesiones diferentes, por otro lado está la **Ahumada**, que recoge la voz de cien locutores varones en cuatro sesiones telefónicas y cuatro sesiones realizadas utilizando el micrófono. Además, esta base de datos está siendo ampliada en el **Diac** con cien mujeres en idénticas condiciones, conformando un grupo de 200 usuarios.

A continuación realizaremos una breve descripción de dos bases de datos para tener una idea de cómo se realizan las grabaciones y bajo que condiciones, y que en cierta forma fueron de gran utilidad desde el punto de vista teórico para la construcción de nuestra base de datos. [10]

3.2.1 Telvoice

Esta base de datos fue realizada en 1998 y fue grabada de forma automática, mediante un ordenador personal, conectado a la línea telefónica, que se encarga de atender la llamada y dirigir el diálogo conducente a la grabación de las distintas frases que se le solicitaban al locutor que llamaba. La frecuencia de muestreo utilizada fue de 8 KHz., codificando cada muestra con 8 bits. Cada locutor debía realizar diez llamadas espaciadas al menos dos o tres días.

El propósito de dicha base de datos, es evaluar los efectos del entrenamiento multisesión y el uso de diferentes técnicas para compensar la variabilidad del canal.

Es de vital importancia considerar en nuestro trabajo, por lo menos desde el punto de vista teórico, la metodología utilizada en ésta base de datos, ya que se toma en cuenta los efectos del canal telefónico.[9]

3.2.2 Diac – 2

Esta base de datos, fue grabada en 1994 con el objetivo de evaluar diferentes técnicas de modelado del locutor, considerar para habla independiente de texto. Está formada por habla procedente de 25 locutores varones. Esto limita en parte la amplitud de la base de datos, al no disponer de datos de hablantes femeninos.

Se trata de una base de datos grabada en una única sesión por locutor (monosesión), por lo que no incorpora, el efecto de variabilidad temporal del hablante. La señal es adquirida a través de un conversor análogo-digital con ancho de banda telefónico, frecuencia de muestreo de 8 KHz. y codificando cada muestra con 8 bits, por lo que a pesar de tratarse de habla recogida por un micrófono, hablaremos de señal con calidad telefónica.

De esta forma, la base de datos incluye entre minuto y medio a dos minutos de habla por cada uno de los locutores con calidad telefónica, lo que va a ser mas que suficiente para evaluar las capacidades del modelado de diferentes sistemas independientes de texto.[9]

3.3 Conceptos Básicos

En primer lugar, es necesario tomar en consideración una terminología adecuada que permita entender la idea de éste proyecto. El campo de investigación está centrado en la Fonética, y más concretamente en el área de la Fonética – Acústica.

La Fonética es el estudio de todos los posibles sonidos que conforman el habla; la Fonología estudia la manera en que los hablantes de una lengua sistemáticamente utilizan una selección de estos sonidos del habla para expresar significado.

Se ha dicho algunas veces que la fonética acústica existió desde la antigüedad, ya que la descripción que del sonido articulado hicieron los gramáticos griegos se basa sobre todo en la impresión auditiva que les causaba la percepción. Creemos que es necesario hacer una distinción, tan necesaria como imprescindible, entre Fonética Acústica y Fonética Auditiva.

La **Fonética Acústica** deberá ocuparse de estudiar los componentes que conforman la onda sonora compleja de los sonidos articulados, y de buscar cuál o cuáles de ellos son imprescindibles para su reconocimiento. Por otro lado, la Fonética Auditiva se interesa por la percepción del sonido con toda la problemática que esto lleva consigo.

Existe también la **Fonética Articulatoria** o genética que ha sido la única utilizada durante mucho tiempo, y la que aún hoy se utiliza en las descripciones de lenguas, sobre todo con un carácter pedagógico. En Fonética Articulatoria la lengua es el órgano más importante de la boca o cavidad oral, ya que gracias a sus posiciones (plana, cóncava o convexa) y sus movimientos hacia arriba y hacia abajo (eje vertical), y hacia adelante y hacia atrás (eje horizontal) se forman cavidades o resonadores, de distintas formas y volúmenes, que son los generadores de los diversos sonidos.

Ahora bien, la existencia de dos aspectos de la misma disciplina, si prescindimos del auditivo, se plantea inmediatamente la idea de saber cuál de

ellos es el más adecuado para la descripción de los sonidos del lenguaje para el establecimiento de los rasgos distintivos de los fonemas. Esta claro, que la aceptación de uno no implica la negación del otro, pues los dos se complementan, además permite aseverar que la Fonética Articulatoria es el origen de la Fonética Acústica.[20]

Comparemos algunos aspectos entre ambas fonéticas:

- ✓ Ambas fonéticas están, en definitiva, involucradas en un fenómeno más amplio que es el proceso de comunicación. En él, lo que importa es la identificación de los fonemas por parte del oyente. Pero, si es cierto que el aspecto acústico en la comunicación parece ser el más importante, no es menos cierto que los hábitos motrices articulatorios desempeñan un papel importante en la identificación lingüística de la onda acústica recibida.

- ✓ Para la Fonética Articulatoria, cada posición de los órganos articulatorios daba origen a un sonido determinado, y la más leve modificación de ése estado originaba uno nuevo. Pero la aparición de la Fonética Acústica en el campo de la Fonética Articulatoria, ha puesto en conocimiento la inexactitud de esos criterios.

- ✓ Finalmente, podemos decir que los datos proporcionados por los análisis acústicos son más objetivos, más fáciles de manejar y menos numerosos que los articulatorios.

Por ello, los datos proporcionados por la Fonética Acústica, son más objetivos, parciales, adecuados y más constantes que los de la Fonética Articulatoria, sin ser una más o menos importante que la otra.[20]

4 MARCO TEÓRICO

4.1 Voz Humana

4.1.1 Fisiología de la producción del habla

Tradicionalmente necesitamos de tres grupos o sistemas de órganos para que, coordinadamente podamos producir lenguaje hablado. Estos grupos son:[22]

- ✓ *Sistema Respiratorio*
- ✓ *Sistema Fonatorio*
- ✓ *Sistema Articulatorio*

El ***Sistema Respiratorio*** comprende los pulmones y los músculos por los cuales se comprimen o dilatan; los tubos bronquiales y la tráquea. Su función principal es la respiración, asegurando el abastecimiento de oxígeno a la sangre.

El ***Sistema Fonatorio*** está compuesto por la laringe, órgano responsable de la producción básica del sonido. Su función principal es la de actuar como válvula de cierre de la vía que conecta hacia los pulmones, evitando que pasen los alimentos a través de la tráquea.

El ***Sistema Articulatorio*** está formado por las estructuras oro-faciales las que se subdividen en nasales y orales. Cumplen principalmente dos funciones:

- ✓ Permitir el ingreso y egreso del aire hacia y desde las cavidades pulmonares (respiración).
- ✓ Y en el caso de las estructuras orales, la ingesta de alimentos.

El habla, como señal acústica, se produce a partir de las ondas de presión que salen de la boca y las fosas nasales de una persona. El proceso comienza con la generación de la energía suficiente (flujo de aire) en los pulmones, la modificación de ese flujo de aire en las cuerdas vocales, y su posterior

perturbación por algunas constricciones y configuraciones de los órganos superiores. Así, en el proceso fonador intervienen distintos órganos a lo largo del llamado tracto vocal, que en nuestro caso asumiremos que se restringe a la zona comprendida entre las cuerdas vocales y las aberturas finales: los labios y las fosas nasales.

El conjunto de órganos que intervienen en la fonación (figura 4.1) pueden dividirse en tres grupos bastante bien delimitados:[19]

- ✓ Cavidades infraglóticas (sistema sub-glotal) u órgano respiratorio
- ✓ Cavidad laríngea u órgano fonador
- ✓ Cavidades supraglóticas.

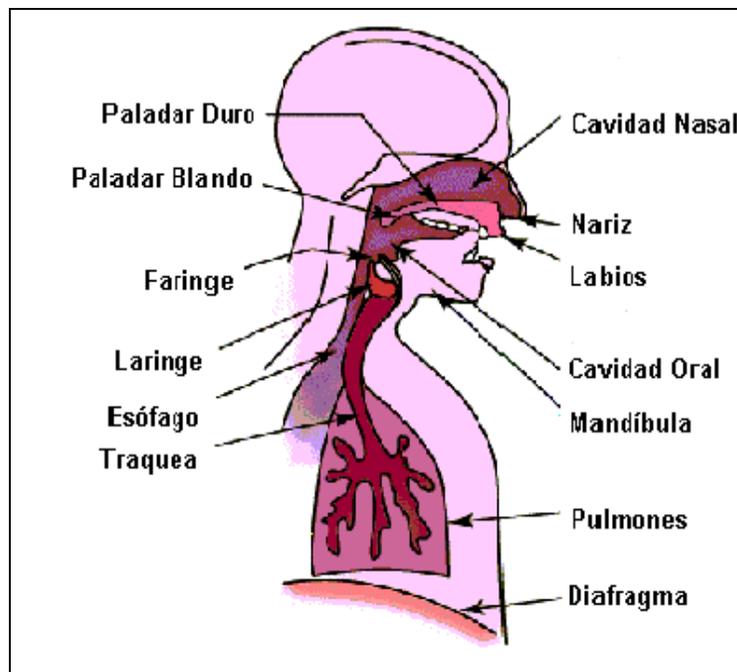


Fig. 4.1- Conjunto de órganos que intervienen en la Fonación. (Fuente: [22]).

4.1.1.1 Cavidades infraglólicas

Las cavidades infraglólicas constan de los órganos propios de la respiración (pulmones, bronquios, y tráquea), que son la fuente de energía para todo el proceso de producción de voz.

En el proceso de inspiración, los pulmones toman aire, bajando el diafragma y agrandando la cavidad torácica. En el momento de la fonación, la espiración, provocada por la contracción de los músculos intercostales y del diafragma, aporta la energía necesaria para generar la onda de presión acústica que atravesará los órganos fonadores superiores.

4.1.1.2 Cavidad laríngea

La cavidad laríngea es la responsable de modificar el flujo de aire generado por los pulmones y convertirlo (o no), en una señal susceptible de excitar adecuadamente las posibles configuraciones de las cavidades supraglólicas.

El último cartílago de la tráquea, el cricoides, forma la base de la laringe, cuyo principal órgano son las cuerdas vocales que son dos pares de repliegues compuestos de ligamentos y músculos. El par inferior son las llamadas cuerdas vocales verdaderas, que pueden juntarse o separarse mediante la acción de los músculos crico-aritenoides lateral y posterior, y que están protegidas en su parte anterior por el cartílago tiroides, el más importante de la laringe, abierto por su parte posterior. Finalmente, la parte superior de la laringe está unida al hueso hioides.

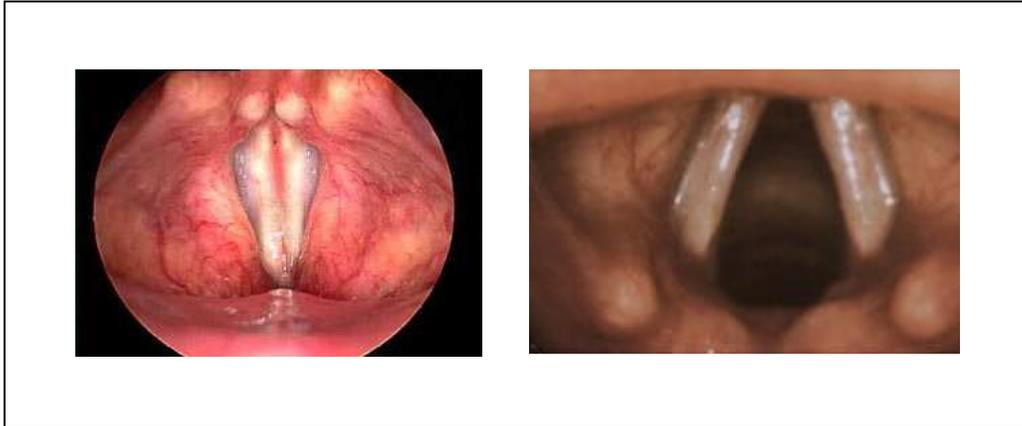


Fig. 4.2. - Vista longitudinal de las cuerdas vocales cerrada y abierta. (Fuente: [12]).

En la figura 4.2 se muestra una vista longitudinal simplificada de la zona en la que se encuentran las cuerdas vocales, en sus posiciones extremas: abiertas y cerradas. A la apertura que queda entre las cuerdas vocales se les denomina glotis. La cavidad laríngea está terminada por la epiglotis, un cartílago en forma de cuchara, que permite cerrar la apertura de la laringe en el acto de la deglución.

4.1.1.3 Cavidades supraglóticas

Las cavidades supraglóticas están constituidas por la faringe, la cavidad bucal y la cavidad nasal. Su misión fundamental de cara a la fonación es perturbar adecuadamente el flujo de aire procedente de la laringe, para dar lugar finalmente a la señal acústica generada a la salida de la nariz y la boca. La faringe es una cavidad en forma tubular que une la laringe con las cavidades bucal y nasal, y que suele dividirse en tres partes: faringe laríngea, faringe bucal (boca) y faringe nasal, las dos últimas separadas por el velo del paladar. El volumen de la faringe laríngea puede ser modificado por los movimientos de la laringe, la lengua y la epiglotis mientras que el volumen de la faringe bucal se modifica por el movimiento de la lengua.

La faringe nasal y las restantes cavidades nasales forman, desde el punto de vista de su acción sobre el flujo de aire procedente de la faringe, un resonador que puede o no conectarse al resonador bucal mediante la acción del velo del paladar. Según el resonador nasal esté o no conectado, el sonido será nasal u oral, respectivamente. Si hacemos una descripción de la cavidad bucal (esquemática en la figura 4.3), podemos señalar las siguientes partes:

- ✓ Los labios en el extremo
- ✓ Los dientes
- ✓ La zona alveolar, entre los dientes y el paladar duro
- ✓ El paladar, en el que a su vez, y de forma simplificada, podemos distinguir el paladar duro y el paladar blando o velo.

La raíz de la lengua forma la pared frontal de la faringe laríngea, y sus movimientos le permiten modificar la sección de la cavidad bucal (movimiento vertical), adelantar o retrasar su posición frente a la de reposo (movimiento horizontal), así como poner en contacto su ápice o la parte trasera con alguna zona del paladar.

El movimiento de los labios también interviene en la articulación, pudiendo ser de apertura o cierre y de protuberancia, alargando en este último caso la cavidad bucal.

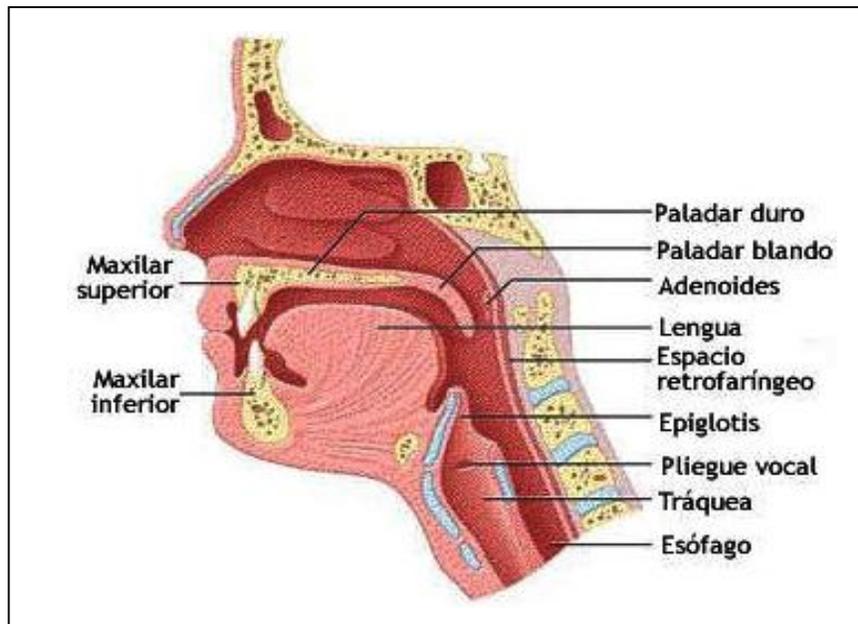


Fig 4.3 - Sección vertical de la cavidad oral. (Fuente: [12]).

4.1.2 Modelo general de producción de la voz

El tracto vocal modelado se manifiesta como un filtro variable, cuyos parámetros varían en el tiempo en función de la acción consciente que se realiza al pronunciar una palabra. El filtro variable en el tiempo tiene dos posibles señales de entrada que dependerán del tipo de señal, sonora o no sonora. Para señales sonoras la excitación será un tren de impulsos de frecuencia controlada, mientras que para las señales no sonoras la excitación será ruido aleatorio. La combinación de éstas señales modela el funcionamiento de la glotis.

El espectro de frecuencias de la señal de la voz puede obtenerse a partir del producto del espectro de la excitación por la respuesta en frecuencia del filtro. El tracto vocal manifiesta un número muy grande de resonancias, sin embargo se consideran solo las tres o cuatro primeras que toman el nombre de **formantes** y cubren un rango de frecuencias entre 100 y 3500 Hz. Esto debido a que las

resonancias de alta frecuencia son atenuadas por la característica del tracto, el cual tiende a actuar como un filtro pasa-bajo con una caída de aproximadamente -12 dB por octava.[14]

Desde un punto de vista mecano acústico, y además considerando que la voz es fisiológicamente normal, las vocales son los sonidos emitidos por la sola vibración de las cuerdas vocales sin ningún obstáculo entre la laringe y las aberturas oral y nasal. Dicha vibración se genera por el principio del oscilador de relajación, donde interviene una fuente de energía constante en la forma de un flujo de aire proveniente de los pulmones. Son siempre sonidos de carácter sonoro y por consiguiente de espectro discreto.

Las consonantes, por el contrario, se emiten interponiendo alguna constricción formado por los elementos articulatorios. Los sonidos correspondientes a las consonantes pueden ser sonoros o no, dependiendo si las cuerdas vocales están vibrando. Funcionalmente, en el español las vocales pueden constituir palabras completas, no así las consonantes.[14]

El reconocimiento de una consonante a través de su percepción depende esencialmente de la presencia de un cambio de frecuencias en sus elementos acústicos consecutivos, mientras que el de una vocal depende de la estabilidad en la frecuencia. Todos los cambios apreciables en la frecuencia de los formantes, excepto aquellos que aparecen en la unión de dos vocales contiguas, constituyen a la percepción de las consonantes; un cambio no apreciable en la frecuencia de las formantes contribuye a la percepción de las vocales. Todas las consonantes necesitan de otros parámetros (transiciones) para ser percibidas claramente. Ahora bien, cuando un espectrograma presenta cambios en las formantes contribuyen a la identificación de las consonantes, mientras que las que presentan una relativa estabilidad en la frecuencia de las formantes, incluso durante un tiempo breve, se identifican como vocales. Los sonidos que funcionan como una consonante en un caso y como vocal en otro no son fonéticamente los mismos, tienen algo en común pero son diferentes, se distinguen entre sí por un rasgo

fonético marcado, es decir, uno se percibe por medio de un cambio en la frecuencia de los formantes, el otro no.

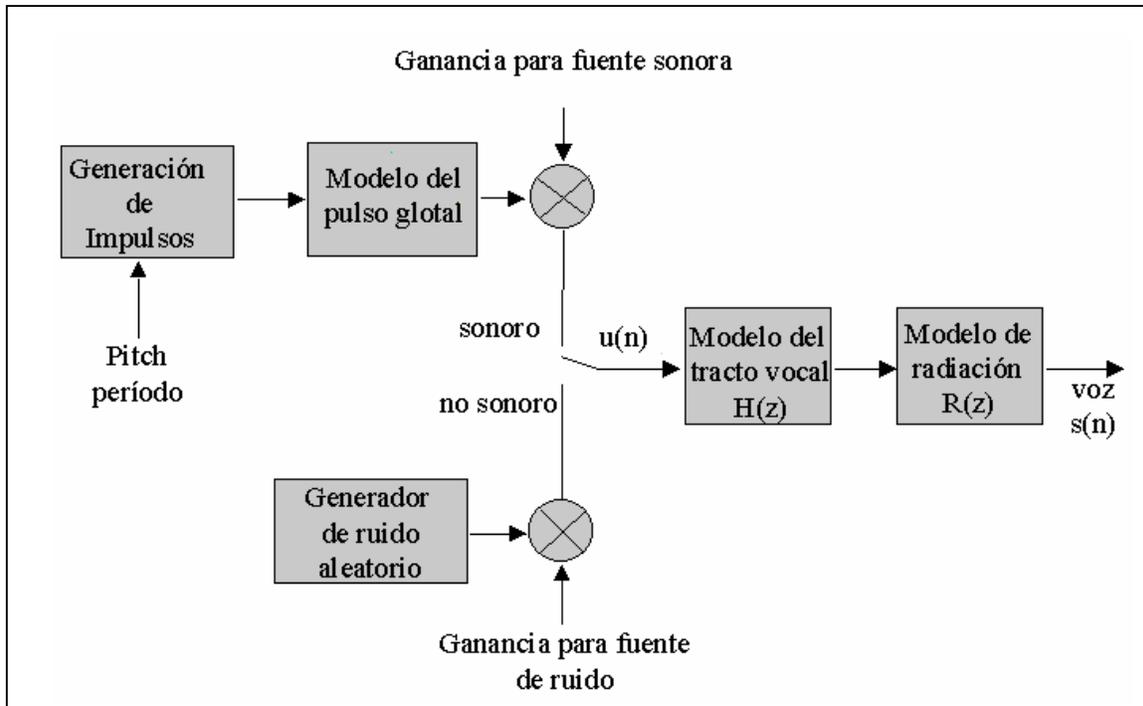


Fig. 4.4 - Modelo general de producción de la voz. (Fuente: [3]).

Este modelo (figura 4.4) es una simplificación del proceso general de producción del habla. Éste modelo supone que las dos señales (señales de frecuencia alta y señales sonoras) pueden interactuar entre ellas sin sufrir ninguna independencia, lo cual es cierto ya que la vibración de cuerdas vocales es afectada por las ondas de presión dentro del tracto.

Si consideramos las señales sonoras que son filtradas en el tracto y las señales no sonoras que provienen del generador de ruido aleatorio, éste modelo supone igualdad en la extensión del filtro para ambas señales, por lo que lo hace más preciso y más confiable a la hora de ser considerado como tal.

Por último, para éste modelo digital de producción de la voz, es necesario considerar la radiación de salida por la importancia que significa la impedancia de radiación al momento de la emisión de la voz.[3]

4.1.3 Modelo acústico de generación de las vocales

En el proceso de generación de la voz humana, específicamente la vocales se pueden identificar tres grandes etapas:[24]

- ✓ *Generación de la fuente de producción*
- ✓ *Modulación del tracto vocal*
- ✓ *Radiación de salida.*

4.1.3.1 Generación de la fuente de producción

En la generación del sonido, se pueden distinguir tres tipos de fuentes de producción. Una es la fuente que produce los sonidos sonoros: se originan gracias a los impulsos de la corriente de aire al pasar a través de la glotis puesta en vibración. Otra es la que produce ruido como un efecto secundario de la turbulenta corriente de aire al pasar a través de constricciones. Otra fuente de producción es la que ocurre cuando se libera el aire comprimido por una obstrucción del conducto vocal.

Sin embargo, si consideramos la fuente que produce los sonidos sonoros, se pueden determinar tres tipos de vibraciones distintas en las cuerdas vocales:

- ✓ **Normal:** Sonido similar a una señal *Diente de Sierra*. Tiene fundamental y todos sus armónicos, con una pendiente de aproximadamente -6dB/octava, sobre 1 KHz.

- ✓ **Cuerdas Separadas (semi-abducidas):** Sonido con turbulencias originadas en el flujo a nivel de las cuerdas. Fundamental y pocos armónicos acompañadas por ruido de banda ancha.
- ✓ **Cuerdas Juntas (abducidas):** Sonido generado por golpes de presión. Baja la frecuencia fundamental y armónicos, acompañados por ruido entre ellos.

En la tabla 4.1 se muestran los rangos comunes de la Frecuencia Fundamental para el Español.

Adulto Hombre	80 - 160 Hz
Adulto Mujer	170 - 340 Hz
Niños	250 - 500 Hz

Tabla 4.1 – Rangos comunes de la frecuencia fundamental. (Fuente: [21]).

4.1.3.2 Modulación del tracto vocal

Al hacer un planteamiento, en el que hablamos de una onda de presión que atraviesa unos espacios de determinada sección (área), es fácil ver cómo es posible modelar el tracto vocal como un tubo, o la concatenación de tubos. Por supuesto, también habría que considerar la cavidad nasal, que siguiendo otros modelos, podría representarse como un tubo en paralelo con la cavidad bucal. Todos los modelos de producción de habla asumen que las cuerdas vocales son independientes del tracto vocal, y que éste último se comporta como un sistema lineal, lo que supone una aproximación bastante razonable de cara a la simplicidad de su tratamiento. Sea como sea, el sistema de producción de habla puede verse como una excitación que atraviesa un canal, y cómo ese canal

modifica las características espectrales de la excitación original. No es el objetivo de este proyecto entrar en los detalles matemáticos del modelo basado en tubos del tracto vocal, con lo que nos limitaremos a comentar brevemente los resultados más importantes.

Si suponemos un modelo simple de un único tubo, cerrado en el punto donde se aplica la excitación (cuerdas vocales), tendremos la situación en el que $u(x, t)$ es la velocidad volumétrica a una distancia x del origen de la excitación en el tiempo t , y en el que asumimos una excitación exponencial (figura 4.5).[18]

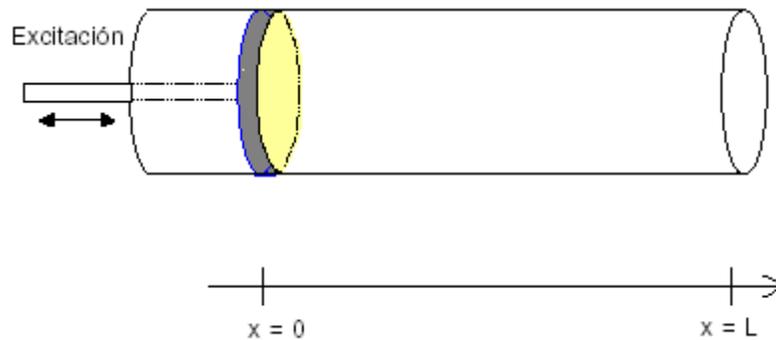


Fig. 4.5 - Modelo simple de un tubo de sección constante. (Fuente: [18]).

Donde tenemos por definición, las velocidades de volumen de los labios y en la glotis respectivamente.

$$u(0,t) = \text{Re}\{U(0)e^{j\omega t}\} \quad (1)$$

$$u(l,t) = \text{Re}\{U(l)e^{j\omega t}\} \quad (2)$$

Partiendo de la ecuación de onda y aplicando las condiciones de contorno adecuadas, podemos llegar a calcular la función de transferencia del sistema (ya en dominio frecuencial y siendo C_0 la velocidad de transmisión del sonido en el medio):

$$T(f) = \frac{U(l)}{U(0)} = \frac{1}{\cos\left[\frac{\omega}{C_0} l\right]} \quad (3)$$

Donde $T(f)$: representa la razón de amplitudes complejas de la velocidad de volumen en los labios con la velocidad de volumen en la glotis.

Con lo que es fácil ver cómo las frecuencias de resonancia (aquellas donde la función de transferencia se hace infinita en ausencia de excitación) ocurren para:

$$\frac{\omega}{C_0} l = \frac{\pi}{2} (2n + 1), \quad n = 0, 1, 2, \dots \quad (4)$$

o lo que es lo mismo para:

$$f = \frac{C_0}{4l} (2n + 1), \quad n = 0, 1, 2, \dots \quad (5)$$

Lo que implica que las resonancias ocurren en múltiplos impares de la frecuencia de resonancia fundamental, $C_0 / 4L$.

Con estos datos, y para un tracto vocal típico de unos 17 cm. de longitud (valor típico para un hombre adulto), las frecuencias de resonancia aparecen en, aproximadamente, 500 Hz., 1500 Hz., 2500 Hz. etc., como se muestra en la figura 4.6.

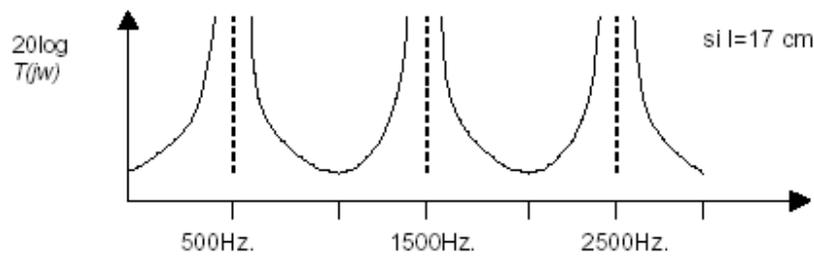


Fig. 4.6 - Frecuencias de resonancia. (Fuente: [18]).

Un modelo más razonable tendría que tener en cuenta muchas más secciones, cada una con un área determinada. Del análisis de modelos de varios tubos de secciones distintas aparece un nuevo concepto: el de coeficientes de reflexión, que nos dan idea de que en la conexión de los tubos hay una porción de la onda que se transmite y otra que se refleja. Los estudios típicos discretizan los cálculos haciéndolos únicamente en las fronteras entre dos tubos, lo que da lugar a modelos de flujos de señal, similares a los utilizados en los estudios sobre guías de onda. Las simulaciones de guías de onda que siguen estos modelos muestran la presencia de múltiples resonancias, que se corresponden con las observaciones experimentales sobre las características espectrales de la señal de voz.

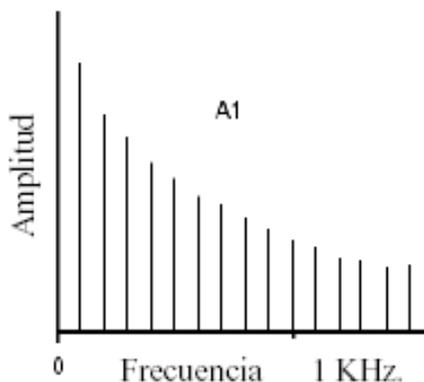


Fig. 4.7a - Espectro de la señal glotal

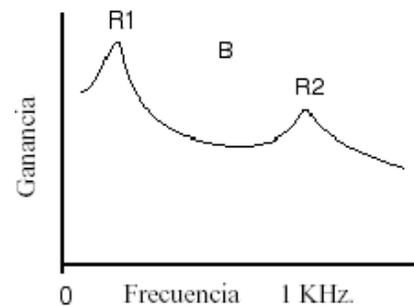


Fig. 4.7b - Función de transferencia del tracto vocal

Así, con el adecuado modelo matemático, seríamos capaces de explicar las medidas experimentales del espectro de la señal de voz real. Veamos resumidamente los conceptos fundamentales: En la figura 4.7a, se muestra el espectro típico de una señal glotal (cuando vibran las cuerdas vocales) donde puede observarse la estructura periódica de la señal (aparece el primer armónico y sus múltiplos, con una caída en amplitud de alrededor de unos 6 dB/octava), donde la separación entre cada dos máximos consecutivos coincide con la frecuencia fundamental de la voz. En la figura 4.7b, se muestra un ejemplo de espectro de la función de transferencia del tracto vocal. El efecto de dicho tracto vocal será modificar el contenido espectral de la señal de voz producida por la combinación, lo que en frecuencia equivale a la multiplicación de los espectros originales, tal y como aparece en la figura 4.8.

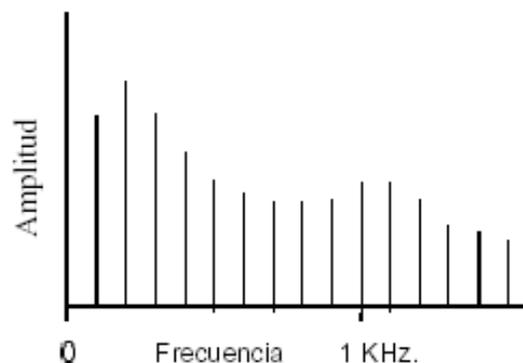


Fig. 4.8 - Señal de la voz después del tracto vocal. (Fuente: [18]).

OBS: Existe un método llamado **“método de perturbación”** mas simple de entender que tiene relación con los resonadores complejos. La interpretación de éste método se basa en el comportamiento individual de cada modo normal de vibración. Si un estrechamiento se produce en una zona cercana a un antinodo, la frecuencia de dicho modo bajará. En caso de ubicarse cercano a un nodo, la

frecuencia de dicho modo aumentará. De modo contrario ocurre este fenómeno para el caso de un ensanchamiento.

Para el caso de un resonador acoplado al sistema como se muestra en la figura, este altera el comportamiento original, agregando una resonancia y una antiresonancia (llamados también polo y cero) y subiendo además la frecuencia de la primera resonancia del sistema original. Estos efectos se hacen más notorios mientras mayor sea la superficie que une ambos resonadores.[24]

Ahora bien, como dijimos anteriormente, a los máximos espectrales generados por las resonancias del tracto vocal se les llama **formantes**, y están en posiciones determinadas para cada uno de los sonidos sonoros que podemos producir, aunque hay una variabilidad inherente en la producción de los mismos por parte de cada persona. Desde el punto de vista del control de la fonación, la persona puede obviamente modificar los elementos de su tracto vocal (movimiento de la lengua, labios, control de la apertura de la cavidad nasal, etc.), y también puede modificar las características de la fuente glotal de dos formas distintas:

- ✓ Variando la frecuencia de vibración de las cuerdas vocales.
- ✓ Variando la forma de onda de la velocidad volumétrica glotal.

La articulación es una modificación principalmente a nivel temporal de los sonidos, y está directamente relacionada con la emisión de los mismos y con los fenómenos transitorios que los acompañan. Está caracterizada por el lugar del tracto vocal en que tiene lugar, por los elementos que intervienen y por el modo en que se produce, factores que dan origen a una clasificación fonética de los sonidos que veremos más adelante.[18]

4.1.3.3 Radiación de salida

La radiación de salida, o la forma de salir que tiene el sonido de los labios y apertura en el tracto vocal, tiene asociada una impedancia de radiación que influye sobre el sonido de varias formas. En primer lugar, la configuración de los labios influye significativamente a la frecuencia de corte del filtro pasa bajos que estos producen, alterando gran parte de los componentes de media y altas frecuencias (sobre 1 KHz.) cuando estos están muy cerrados. Por ejemplo, al comparar las vocales “u” y “a” de la lengua española, no solo cambia las formantes, sino también la frecuencia de corte de filtro de radiación. Esto influye en la energía total irradiada y en la capacidad de emitir eficientemente la energía vocal (proyección de la voz).

Se pueden hacer las siguientes observaciones considerando el patrón polar de la señal de la voz (figura 4.9).[24]

- ✓ En baja frecuencia el efecto de la cabeza es despreciable.
- ✓ En frecuencias medias varía en aproximadamente 1 dB en 90° con respecto a 0°.
- ✓ En frecuencias altas varía en aprox. 5 dB en 90° con respecto a 0°.

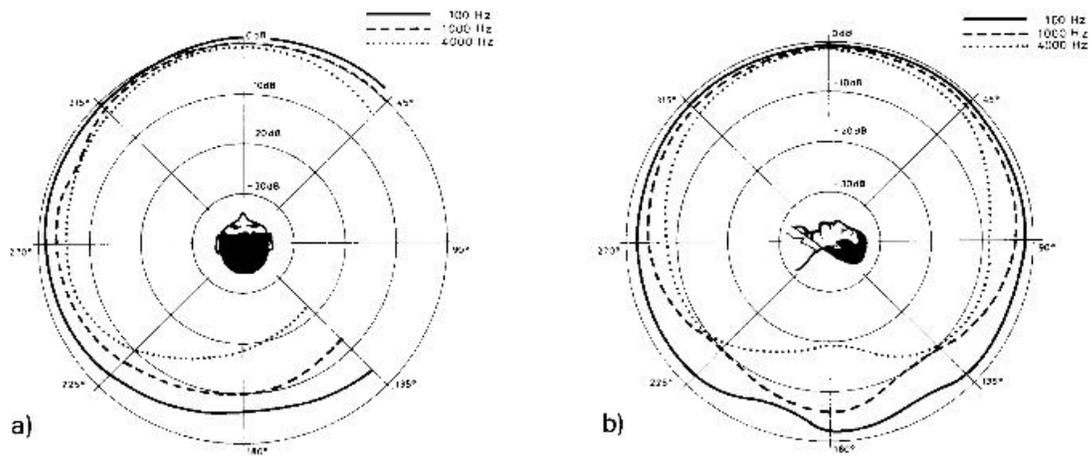


Fig. 4.9 - Patrón Direccional de la voz. En la Figura a) se describe el patrón direccional en posición vertical. La figura b) muestra el patrón direccional en posición horizontal. (Fuente: [24]).

4.1.4 Propiedades de las Formantes

Las formantes sufren variaciones con respecto a las amplitudes del espectro del sonido generado en la fuente de producción, pero aún así, no desplazan las frecuencias propias del sistema original.

Ahora bien, si observamos el fenómeno de resonancia producido en el tracto bucal producto de las variaciones de las secciones del mismo, se puede asumir que varían las frecuencias de las formantes y además todos los formantes tienen un ancho de banda definido:[24]

$$B(fn) = 50(1 + fn/1000) \quad (6)$$

Donde (fn) : representa la frecuencia formántica, con $n = 0, 1, 2, \dots, n$

Es importante señalar que tanto para los sonidos sonoros vocálicos y consonánticos los análisis normalmente se realizan considerando los cuatro

primeros formantes, donde el primer formante generalmente controla la amplitud del sonido, y el segundo por las variaciones frecuenciales que posee controla generalmente la inteligibilidad del sonido.

4.1.5 Fonemas vocálicos y consonánticos

La mínima unidad lingüística capaz de producir cambios de significados se le llama **fonema**. Tanto los fonemas vocálicos como los fonemas consonánticos pueden ser clasificados según el estado vibratorio de las cuerdas vocales en fonemas fonados (ó sonoros) y áfonos (ó sordos).

Los **fonemas fonados** se producen por la acción de las cuerdas vocales, que obturan el paso del flujo de aire durante breves instantes de tiempos. Éste tren de pulsos es modificado por los órganos de articulación que vienen posteriormente añadiendo una cierta codificación. Esta codificación consiste en el resalte de determinadas frecuencias contenidas en el pulso glotal original o en su eliminación. Por lo tanto el sonido así emitido tiene una forma relativamente periódica.

Los **fonemas áfonos** se producen si la acción de las cuerdas vocales se ve alterada (posición de respiración). El flujo de aire procedente de los pulmones se vuelve más rápido, produciendo fricciones y turbulencias. Esto se traduce en múltiples vibraciones en aquellos puntos en que por producirse un estrechamiento, la velocidad del flujo de aire alcanza sus máximos. Ahora bien, si el punto de estrechamiento o articulación se encuentra próximo al exterior (labios, dientes), el resto de la estructura influye escasamente en el sonido emitido. En caso contrario (zona velar o medial), la posición de los restantes órganos puede modular bastante el resultado. Su forma de onda presenta un escaso carácter repetitivo, recordando al ruido aleatorio.

Finalmente se pueden decir que la principal diferencia entre los Fonemas Vocálicos y Consonánticos es el tamaño de estas constricciones asociadas a las articulaciones en el tracto, siendo en las consonantes estas obstrucciones mucho

más estrechas que en las vocales y además los Fonemas Vocálicos son siempre fonados, a diferencia de los consonánticos que pueden ser o no fonados.

A continuación en las figuras 4.10 y 4.11 se mostrarán espectrogramas de fonemas fonados y áfonos, es decir vocal **a** y consonante **f** respectivamente.[21]

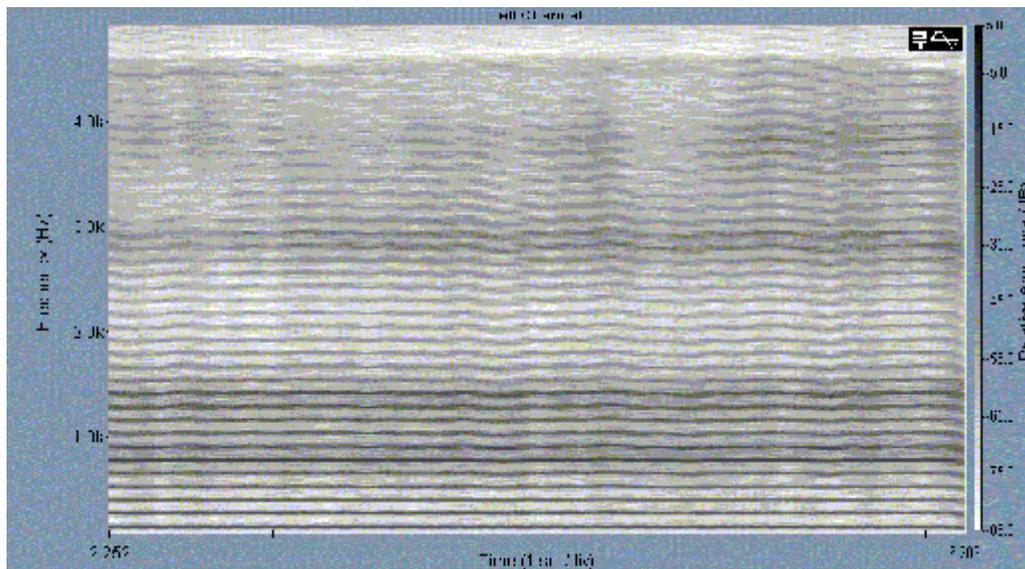


Fig. 4.10 – Espectrograma del fonema fonado /a/, donde se distinguen las tres primeras formantes considerando la ubicación donde existe mayor cantidad de energía. (Fuente: Elaboración propia).

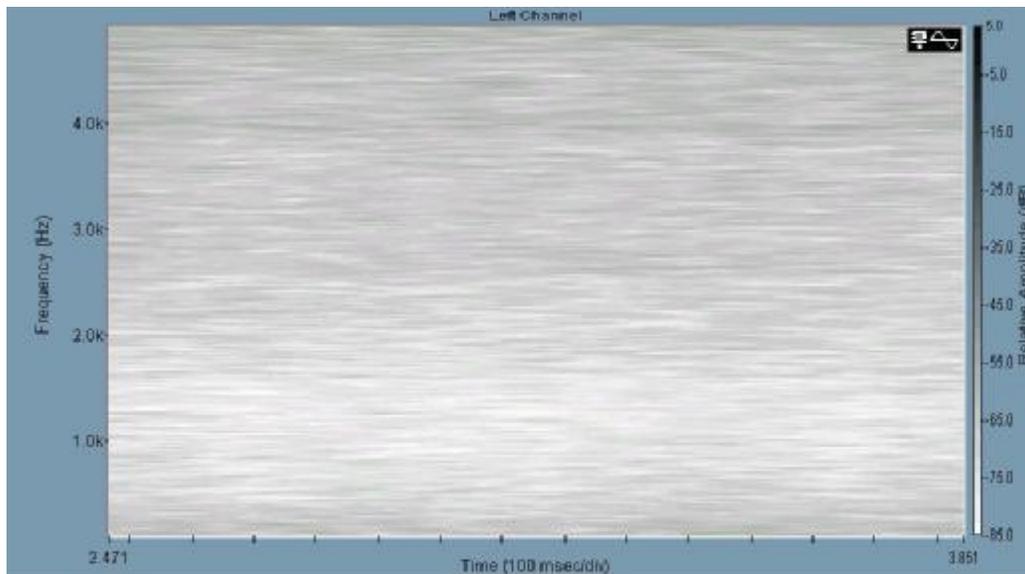


Fig. 4.11 – Espectrograma del fonema áfono /s/, donde no se distinguen formantes por que la concentración de energía es superior a los 7000 Hz. (Fuente: Elaboración propia).

Si realizamos un breve análisis de los espectrogramas expuestos anteriormente, debemos considerar que:

- Fue considerado éste rango por las características propias de una señal telefónica, cuyo ancho de banda (como se detalla con posterioridad) es de 3,5 KHz.
- Si tomamos en cuenta el ancho de banda de la voz, donde por la naturalidad con que se habla y la inteligibilidad con que se percibe oscila entre los 50 Hz. y los 7 KHz.

4.1.6 Producción de vocales

Existen varias teorías con respecto a la producción de las vocales. Una de ellas se relaciona a cada abertura de la glotis, una cierta cantidad de aire penetra en la cavidad bucal, la cual excita el aire de la cavidad en una vibración libre amortiguada; cuando ésta vibración se repita periódicamente e irradia fuera de la cavidad como un sonido, el resultado es una vocal. Otra teoría se asocia a la idea de cómo un tono laríngeo resuena con la cavidad bucal y aquellos de sus tonos parciales armónicos cuyas frecuencias naturales de la cavidad son reforzadas, mientras que las otras son debilitadas; el resultado es una vocal. Ambas teorías no se excluyen, sino se complementan.

Sin embargo, está claro que las vocales son fonemas fonados o sonoros que son capaces de distinguirse entre ellos por la radiación de salida y sus formantes, donde las dos primeras formantes bastan para caracterizar el timbre de las vocales, por lo tanto, si a los fonemas vocálicos se le asocia una articulación, que consiste en la modificación de la acción filtrante de los diversos resonadores, lo cual depende de las posiciones de la lengua (tanto en elevación como en profundidad o avance), de la mandíbula inferior, de los labios y del paladar blando, se pueden clasificar en base a dos parámetros:

- ✓ **Apertura del tracto vocal:** vocal abierta y vocal cerrada.
- ✓ **Grado de elevación del dorso de la lengua:** vocales anteriores, vocales centradas y vocales.

Posición vertical	Tipo de vocal	Posición horizontal (avance)		
		Anterior	Central	Posterior
Alta	Cerrada	i		u
Media	Media	e		o
Baja	Abierta		a	

Fig. 4.12.- Clasificación de los fonemas vocálicos considerando su articulación. (Fuente: [17]).

4.1.6.1 Cartas de formantes

Para representar los valores formánticos que a la larga son los que caracterizan a la vocal es necesario utilizar una escala en la que aparezcan debidamente representados los valores de la primera formante y los de la segunda formantes. Esta comparación de las distintas realizaciones vocálicas se lleva a cabo a través de las cartas de formantes (figura 4.13).[22]

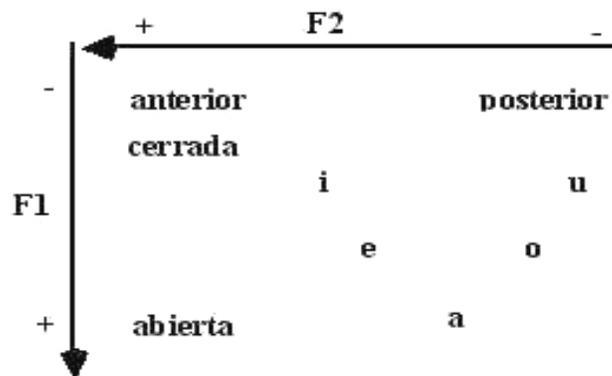


Fig. 4.13 - Carta de formantes. (Fuente: [11]).

Es necesario señalar que la escala utilizada es logarítmica, porque refleja una compensación tal que los intervalos sonoros iguales que llegan a nuestro oído se representan por distancias iguales.

La figura anterior representa una carta de formantes, donde el eje de las abscisas (el horizontal, que va desde los 500 Hz. a los 4000 Hz.) lleva los valores de F_2 ; sobre el eje de las ordenadas (el vertical que va desde los 200 Hz. hasta los 1000 Hz.) se llevan los valores de F_1 .

En el eje de las ordenadas se puede observar el grado de abertura del conducto vocal, es decir, existe una relación constante y directa entre la abertura bucal y el nivel de frecuencias representado en el eje de las ordenadas. Por el contrario, el eje de las abscisas indica la longitud de la cavidad bucal, es decir existe una relación constante e inversa entre la longitud de la cavidad bucal anterior y el nivel de frecuencias representado en el eje de las abscisas.

A modo de ejemplo se puede decir que la vocal [a] aparece en la parte inferior de la carta de formantes, y que tiene, por lo tanto, la frecuencia más alta, posee la mayor abertura del conducto vocal. Lo contrario sucede con la vocal [i], [u].

4.1.6.2 Triángulo vocálico del español

La figura 4.14 representa una carta de formantes sobre un estudio realizado en España, donde se analizaron los cinco fonemas del español en posición fonética normal. En el eje de las abscisas están representados los valores de F_2 y sobre el eje de las ordenadas los valores de F_1 . [21]

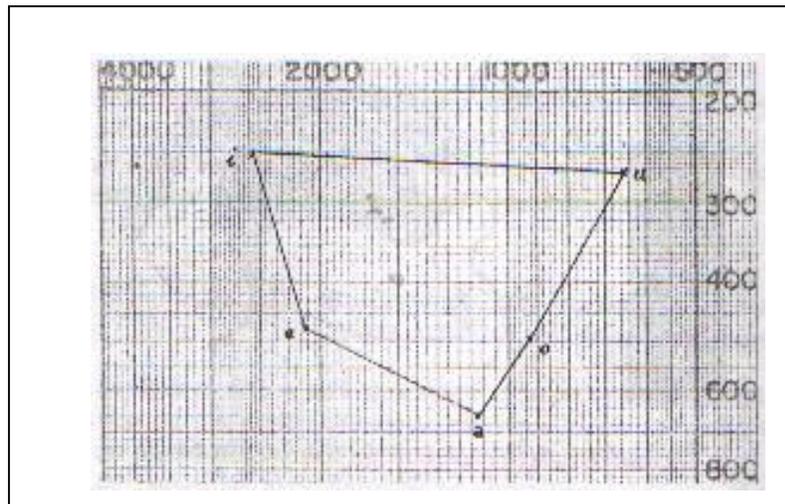


Fig. 4.14 – Triángulo vocálico de la lengua española. (Fuente: [21]).

4.1.7 Producción de consonantes

Las consonantes son fonemas áfonos y en algunos casos fonados. Son producidas con una estrecha constricción en alguna región del flujo de aire arriba de la laringe en donde para las vocales el flujo de aire está más abierto.

Las consonantes pueden ser clasificadas de varias maneras dependiendo de la articulación, que es la responsable de hacer la constricción en el tracto vocal, el grado de constricción, el estado de la glotis y las cuerdas vocales, cuando la constricción se forma, y en donde se forme o no la presión a través de la constricción. Si la constricción de la consonante es suficientemente estrecha existe una caída de la presión a través de la constricción cuando una presión subglotal es aplicada, asumiendo que no hay un bypass en el camino del flujo, así como sucede en los conductos nasales.

Después de la emisión de la consonante, así como los articuladores se mueven hacia configuración apropiadas para la siguiente vocal, las frecuencias de las formantes sufren cambios. Estas transiciones de las formantes proveen información adicional concerniente al lugar en el trato vocal donde la constricción de

las consonantes se ubica.

La transición de los formantes, particularmente F_2 y F_3 son evidentes en el sonido que sigue la emisión de las consonantes, como son los diferentes espectros del ruido.[14]

De acuerdo a la articulación y sus características, podemos clasificarlas en cuatro grandes grupos:[18]

- ✓ *Punto de articulación*
- ✓ *Modo de articulación*
- ✓ *Función de las cuerdas vocales*
- ✓ *Posición del velo del paladar*

4.1.7.1 Punto de articulación

Son los puntos de estrechamiento en las cavidades supraglóticas, antes mencionados, para los sonidos sordos y sonoros, además son controlados por los órganos móviles contra los órganos inmóviles de su misma zona.

Una lista de posibles puntos de articulación puede ser:

- ✓ ***Bilabiales:*** oposición de ambos labios.
- ✓ ***Labiodentales:*** oposición de los dientes superiores con el labio inferior.
- ✓ ***Linguodentales:*** oposición de la punta de la lengua con los dientes superiores.
- ✓ ***Alveolares:*** oposición de la punta de la lengua con la región alveolar.
- ✓ ***Palatales:*** oposición de la lengua con el paladar duro.
- ✓ ***Velares:*** oposición de la parte posterior de la lengua con el paladar blando.
- ✓ ***Glotaes:*** articulación en la propia glotis.

4.1.7.2 Modo de articulación

El modo de articulación define la forma concreta en que la articulación puede tener lugar. Existen dos posibilidades:

- ✓ *modo espirado*
- ✓ *modo no espirado.*

El **modo espirado** se produce bajo un flujo de aire de origen pulmonar y es el modo habitual en la mayoría de las lenguas. El **modo no espirado** se produce por compresión o enrarecimiento del aire en la cavidad bucal, que sufre dos cierres, uno de los cuales cede bruscamente con entrada o salida de flujo aéreo.

Dentro del modo espirado y dependiendo del tipo de aproximación practicada por los órganos articulatorios podemos tener dos submodalidades: **consonantes, sonantes.**

I) Consonantes

Corresponden al máximo cierre del tracto vocal en algún punto concreto. Pueden ser: Oclusivas y Fricativas.

- ✓ **Oclusivas:** Se produce una obstrucción total al paso del flujo de aire, pero solamente durante breves instantes, dando lugar a una explosión posterior a la apertura del tracto vocal. Su duración por lo general es de entre 10 a 30 ms.
- ✓ **Fricativas:** Se produce una obstrucción incompleta del flujo, que busca los pequeños resquicios residuales que puedan existir, aumentando enormemente la velocidad del fluido en la zona. Puede tener una duración superior a los 100 ms.

II) Sonantes

Son aquellas articulaciones en las cuales no se produce estrechamiento suficiente para que aparezca fricción, y sólo se realiza una modificación por cierre parcial o derivación. Pueden ser: Nasales, Laterales, Vibrantes y Aproximantes.

- ✓ **Nasales:** La cavidad nasal se abre al paso del flujo aéreo, produciendo una oclusión total o casi total en algún punto de articulación de la cavidad oral. La radiación es bastante menor en amplitud que la que se produciría por el tracto vocal.
- ✓ **Laterales:** Se basan en una obstrucción incompleta alrededor del tercio anterior de la lengua. El ápice bloquea total o parcialmente al flujo glotal, pero los bordes laterales de la lengua permiten en el paso por el espacio que queda a ambos lados de la lengua.
- ✓ **Africados:** Se define como una oclusión que viene seguida por fricación.
- ✓ **Vibrantes:** Variante de las líquidas, en las que el ápice realiza bruscas aperturas y cierres de su contacto con la zona alveolar o prepalatal, junto con la oclusión total o parcial del flujo lateral por medio de las zonas laterales de la lengua.
- ✓ **Aproximantes:** Son sonidos fuertemente relacionados con la evolución de un sonido a otro en la articulación de un diptongo. En tal caso, uno de los sonidos de tipo vocálico se consonantiza por cierre excesivo. Para su articulación requieren siempre de la presencia de una vocal.

Los fonemas Oclusivos (correspondientes a las consonantes “b”, “c”, “k”, “d”, “g”, “p”, “t”) también se denominan a veces explosivos, debido a la liberación repentina de la presión presente inmediatamente antes de su emisión. Pueden ser sordos o sonoros, al igual que los Fricativos (“b” intervocálica, “f”, “j”, “h” aspirada, “s”, “y”, “z”). Sólo existe un fonema Africado en español, correspondiente a la “ch”.

Los laterales (“l”, “ll”) a veces se denominan líquidos, y son siempre sonoros. Los dos fonemas vibrantes del español (consonantes “r”, “rr”) difieren en que en uno de ellos (“r”) se ejecuta una sola vibración y es intervocálico, mientras que en el otro (“rr”) es una sucesión de dos o tres vibraciones de la lengua. Finalmente, los fonemas aproximantes (la “i” y la “u” cerradas que aparecen en algunos diptongos) son a veces denominados semivocales, pues en realidad suenan como vocales. Pero exhiben una diferencia muy importante: son de corta duración y no son prolongables.

En la figura 4.15 se indican las consonantes clasificadas según el lugar y el modo de articulación, la sonoridad y la oro-nasalidad. En algunos casos una misma consonante aparece en dos categorías diferentes, correspondiente a las diferencias observadas.

Lugar de articulación	Modo de articulación								
	Oral							Nasal	
	Oclusiva		Fricativa		Africada	Lateral	Vibrante	Aproximante	Sonora
	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	Sonora	Sonora	
Bilabial	p	b, v		b, v				w	m
Labiodental			f						
Linguodental			z						
Alveolar	t	d	s	y	ch	l	r, rr		n
Palatal				(y)	(ch)	ll		i	ñ
Velar	k	g	j						
Glotal			h						

Fig. 4.15 - Clasificación de los fonemas consonánticos según su lugar y modo de articulación. (Fuente: [18]).

4.1.7.3 Función de las cuerdas vocales

Se asocia a la idea de si hay presencia o ausencia de las cuerdas vocales al momento de producir un fonema.

4.1.7.4 Posición del velo del paladar

Está relacionado con la capacidad de controlar el paso de aire hacia la cavidad nasal u oral.

Un ejemplo de una fonema fricativo /s/, donde el flujo de aire a través de la constricción es dirigida en contra de los incisivos bajos, y la fuente principal del ruido de la turbulencia está localizada alrededor de los dientes.

La generación del sonido para este fonema puede ser modelado por un tubo constrictivo con un estrechamiento (figura 4.16). La fuente del ruido está representada como una fuente de presión P_s . Las dimensiones típicas de un adulto para la configuración del fonema /s/ son la longitud de la cavidad frontal $L_f = 2$ cm. (incluyendo la corrección para la reactancia de la radiación), distancia desde la constricción al obstáculo 1 cm, y un área seccional cruzada de la cavidad frontal 1 cm. La magnitud de la función de transferencia U_0/P_s , exhibe un peak que corresponde a la resonancia de la cavidad frontal que se encuentra alrededor de los 4400Hz. [14]

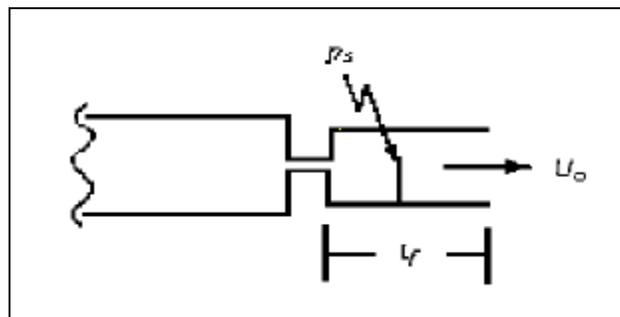


Fig. 4.16 - Modelo acústico de la generación del fonema fricativo /s/. (Fuente: [17]).

4.2. Voz Normal

Es difícil decir si una voz es normal y es más difícil aún decir como debería ser una voz para considerarla normal en una persona determinada.

La estructura laríngea y las características de la voz son un fiel reflejo de la edad del individuo, del sexo y del estado de salud. Esta maduración se prolonga a lo largo de los distintos períodos vitales que determinan modificaciones estructurales y fonéticas notables.

4.2.1 Períodos de evolución

- ✓ **Niñez:** Dentro de éste período podemos encontrar tres subdivisiones, tales como
 - § **Neonatal:** se relaciona con las modificaciones fónicas (llanto, grito) y se caracterizan por ser sonidos de altas frecuencias. El ataque del sonido es brusco de fuerte intensidad y de modulación muy reducida.
 - § **Primera Infancia:** en ésta etapa (primer mes de vida) comienza a modificarse la voz, el ataque del sonido se hace menos brusco, coincidiendo con la instalación paulatina de la capacidad funcional hormonal propia del niño.
 - § **Segunda Infancia:** de similares características a la etapa anterior, pero las variaciones vocales llegan hasta una octava y media de extensión. A los 18 meses aparece la modulación vocal.

- ✓ **Pubertad:** Esta etapa se relaciona por la acción de la estimulación hormonal, que condicionan ciertas características propias de cada sexo y de su capacidad funcional. El cambio fisiológico de la voz, se

caracteriza por signos evidentes, cuyas dualidades fónicas dominan el cuadro, es decir pasan a predominar los armónicos y las resonancias pectorales. Es una voz que pasa del grave al agudo produciendo los conocidos “gallos”, existe una disminución en la intensidad de la voz y en algunos casos se hacen presente ciertas disfonías. Estos cambios se les denomina **Puberfonías** y ocurren normalmente entre lo 11 y/o 16 años.

- ✓ **Período de Estabilización:** Principalmente éste periodo consiste en la fase de estabilidad vocal, es decir la época de juventud - adulto hasta aproximadamente los 60 años, considerando tanto los hombres como las mujeres.
- ✓ **Senectud:** Por último en ésta etapa de la voz, podemos decir que se produce una gran pérdida de agudos, una leve disminución en la extensión de los sonidos, una disminución en las frecuencias armónicas y una pérdida de potencia en el sonido.

Estos periodos de evolución son propios de cada individuo, y en ningún caso se consideran factores culturales, sociales o de educación al momento de señalar las características fónicas de la voz en el tiempo.[12]

Existen varias teorías en relación a la normalidad de una voz. Una de ellas plantea que “hay alteración de la voz cuando difiere de las voces de otras personas del mismo sexo y similar edad y grupo cultural en timbre, tono, volumen, flexibilidad en la dicción. El concepto de voz normal, con unos criterios absolutos, no existe”. [13]

Otra teoría dice que “es obvio que no existe una forma única de sonido que podamos llamar voz normal, existiendo voces de niños, de niñas, voces de hombres y de mujer, voces de anciano... Entre éstos grupos además pueden haber voces normales y anormales. El umbral que separa lo uno de lo otro lo juzga

cada observador en base a sus criterios culturales, educativos, ambientales, de conocimiento vocal y factores similares, pero donde quiera que coloquemos la separación entre lo normal y lo patológico es evidente que cada uno tiene ideas adquiridas sobre lo que es normal y sobre lo que es patológico”.

Por lo tanto es posible sólo establecer criterios generales sobre la voz normal basados en lo siguiente:[6]

4.2.2 Criterios de Normalidad

- ✓ El volumen debe ser el apropiado, sin que la voz sea tan débil que no se pueda oír en condiciones de un entorno sonoro normal, ni que sea tan alta que llame negativamente la atención.
- ✓ Desde el punto de vista de la percepción auditiva se pueden rescatar tres ideas:
 - § El tono debe ser el adecuado en correspondencia a la edad y al sexo del individuo.
 - § Voz con resonancia oral para los fonemas orales y con resonancia nasal para los fonemas nasales.
 - § La voz no debe presentar escape de aire.
- ✓ Tono comprendido entre los 120 Hz. y los 230 Hz.
- ✓ Desde el punto de vista de la funcionalidad, la persona debe cumplir con los objetivos comunicacionales del usuario, esto quiere decir que si un locutor, por ejemplo, no posee una voz adecuada para ser locutor, no tiene voz normal, pero si pueda quizás desempeñarse profesionalmente en otra área en condiciones normales.

4.2.3 Tipos de voz

Desde el punto de vista de la sonoridad podemos distinguir al menos 4 tipos de voz.[6]

- ✓ **Voz Estridente:** básicamente se caracteriza por ser una voz de sonoridad extrema producto de una tensión a nivel de la faringolaríngea y es producida principalmente por una técnica vocal deficiente. Espectralmente la voz estridente tiene presencia de ruido sobre los 1000 Hz.
- ✓ **Voz Opaca:** es una voz con escape de aire producto de un cierre incompleto en cuerdas. Este escape de aire puede ser de grado variable o simplemente una vibración anómala de las cuerdas vocales. Tiene una caída en la amplitud en las frecuencias sobre los 2500 Hz.
- ✓ **Voz Clara:** también conocida como voz blanca, se caracteriza por ser una voz con poca riqueza acústica, es decir, la amplitud se ve disminuida en frecuencias hasta los 1000 Hz. Las dimensiones del tracto vocal se ven reducidas pero sin ser consideradas como una tensión en dicha zona.
- ✓ **Voz Oscura:** o voz gruesa, se caracteriza por ser básicamente una voz con gran amplitud en las frecuencias hasta los 1200 Hz. Y las dimensiones del tracto vocal también se ven alteradas y ser consideradas como un efecto tensional.

Finalmente la voz, además de su propia sonoridad, es un puente de transmisión para determinados estados de ánimo en una persona, por ende es de vital importancia tomar en consideración su cuidado y no exigirla en demasía.

4.3 Patologías del Habla

Las patologías del habla, serían aquellas que incapacitan a una persona para articular de forma correcta y emitir mensajes hablados. Si la voz normal es la que no llama la atención en ninguna de sus características, la voz patológica será la que no responda entonces a criterios de similitud para grupos de la misma edad, sexo y educación.

En las voces patológicas lo que se puede definir es quién marca el criterio de similitud o de diferencia; éste criterio puede ser notado por la propia persona, las demás personas que están en su entorno o simplemente un especialista. Si sólo es la propia persona quien nota ésta anomalía, por lo general se puede pensar que no es demasiado importante.[13]

Existen diversos tipos de patologías, las cuales podrían ser clasificadas a modo simple según el elemento de la cadena que fuese alterado:[23]

- ✓ **Patologías a nivel de la fuente de producción:** Glotis, Laringe.
- ✓ **Patologías en los elementos Resonantes:** Tracto Vocal.
- ✓ **Patologías en los elementos de Radiación:** Labios.

La amplitud de la voz puede variar por numerosas razones. Irregularidades sufridas período a período, también pueden ser asociadas con la incapacidad de las cuerdas vocales de soportar una vibración periódica por un período definido y por la presencia de ruido de turbulencia en la señal.

El tono en la voz puede variar por la incapacidad de las cuerdas vocales de soportar una vibración periódica por un periodo definido. Usualmente estos tipos de variaciones son aleatorias y son normalmente asociadas con voces roncadas o afónicas.

Es posible determinar algunos patrones patológicos comparando los sonidos con voces normales. Estos sonidos tienen que cumplir con al menos un par de características, tales como la prolongación en un período de tiempo

considerable (mayor a tres segundos) y la amplitud no debe sufrir cambios significativos. Por lo tanto podemos decir que las patologías de la voz es posible estudiarlas por medio de las vocales.

4.3.1 Descriptores Acústicos

Existen una serie de descriptores acústicos que permiten en cierta forma, y bajo las condiciones explicadas anteriormente, tener un mayor detalle de una voz patológica:[23]

- ✓ Perturbaciones de F_0
- ✓ Perturbaciones de la Amplitud
- ✓ Razón de Ruido Armónico
- ✓ Mediciones del Pitch
- ✓ Quiebres de Voz

A continuación se describirán algunos descriptores acústicos, que tienen mayor relevancia al momento de realizar el análisis de una voz.

4.3.1.1 Perturbación de F_0

Jitter [%]: Evaluación relativa de la variación del tono hecha período a período (en pequeños intervalos) sobre la muestra de voz analizada. El Jitter es una medición relativa y la influencia de la frecuencia fundamental promedio del sujeto es significativamente reducida. Es sensible para variaciones de tono que ocurren entre períodos consecutivos del tono. Sin embargo, erróneas extracciones del tono pueden afectar significativamente el porcentaje de Jitter.

El umbral de normalidad es de 1.04%, y se calcula a partir de la información del tono extraída período a período:[23]

$$Jitter = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_o^{(i)} - T_o^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N T_o^{(i)}} \quad (7)$$

Donde: $To^{(i)}$ [$i = 1, 2, \dots, N$] = Período de Fo ; información del período del tono extraído
 N = Número de períodos del tono extraído.

VFo: Coeficiente de variación de Fo [%] - Es la desviación estándar relativa de la frecuencia fundamental. Expresa la variación de Fo (cortos o largos intervalos) en una muestra de voz analizada. El valor de VFo aumenta independiente del tipo de variación del tono. También se incrementa el valor con variaciones en intervalos aleatorios o regulares. Los umbrales normados de fonación sostenida asumen que la frecuencia fundamental no debe cambiar, cualquier variación de la frecuencia fundamental, se verá reflejado en VFo.

Estos cambios pueden ser temblores de frecuencia o cambios no periódicos, un muy alto Jitter o simplemente una subida o caída del tono fuera de la longitud analizada. Se calcula como la razón entre la desviación estándar de la información de Fo extraída período a período y la frecuencia fundamental promedio:

$$vFo = \frac{s}{Fo} = \frac{\sqrt{\sum_{i=1}^N \left(\frac{1}{N} \sum_{j=1}^N (Fo^{(j)} - Fo^{(i)})^2 \right)}}{\frac{1}{N} \sum_{i=1}^N Fo^{(i)}} \quad (8)$$

Donde: $To^{(i)}$ [$i = 1, 2, \dots, N$] = Período de Fo ; información del período del tono extraído
 N = Número de períodos del tono extraído.

4.3.1.2 Perturbación de la Amplitud

Shimmer [%]: Es la evaluación relativa de la variación en la amplitud peak-to-peak, hecha período a período (en pequeños intervalos) sobre una muestra. Es sensible para variaciones de amplitud que ocurren entre períodos consecutivos del tono. Sin embargo inadecuadas extracciones del tono pueden afectar significativamente el porcentaje de Shimmer. El umbral de normalidad es 3,81%, y se calcula partir de la información de la amplitud peak-to-peak extraída:

$$Shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^{(i)} - A^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^N A^{(i)}} \quad (9)$$

Donde: $A^{(i)}$ [$i = 1, 2, \dots, N$] = Período de F_0 ; información de la amplitud peak-to-peak extraída
 N = Numeración de períodos del tono extraído.

VAm: Coeficiente de variación de la Amplitud [%] - Desviación estándar relativa de la amplitud peak-to-peak. Expresa en general variaciones de la amplitud peak-to-peak (en pequeños intervalos) en la muestra de voz analizada. Es calculado como la razón entre la desviación estándar de la amplitud peak-to-peak y el valor promedio de la información de la amplitud peak-to-peak extraída:

$$vAm = \frac{\sqrt{\frac{1}{N} \sum_{j=1}^{N-1} (A^{(j)} - A^{(i)})^2}}{\frac{1}{N} \sum_{i=1}^N A^{(i)}} \quad (10)$$

Donde: $A^{(i)}$ [$i = 1, 2, \dots, N$] = Período de F_0 ; información de la amplitud peak-to-peak extraída
 N = Número de impulsos extraídos

ShdB: Shimmer en dB [dB] - Evaluación en dB de la variación período a período (en pequeños intervalos) de la amplitud peak-to-peak tomada de la muestra de voz analizada. Los quiebres de voz son excluidos. Es calculado de la información de amplitud peak-to-peak que es extraída:

$$ShdB = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left(A^{(i+1)} / A^{(i)} \right) \right| \quad (11)$$

Donde: $A^{(i)}$ [$i=1,2,...N$]=Período de Fo; información de la amplitud peak-to-peak extraída
 N = Número de impulsos extraídos

4.3.1.3 Medición de Pitch

Fo: Frecuencia fundamental promedio [%] - Valor promedio de todos los valores promedio de la frecuencia fundamental, extraídos período a período, y es calculada a partir de la información del tono extraído.

$$Fo = \frac{1}{N} \sum_{i=1}^N Fo^{(i)} \quad Fo^{(i)} = \frac{1}{To^{(i)}} \quad (12)$$

Donde: $To^{(i)}$ [$i=1,2,...N$]=Período de Fo; información del período del tono extraído
 N = Numeración de períodos del tono extraído.

Flo: Frecuencia fundamental mínima o inferior [Hz] - Es el valor mínimo de todas las frecuencias fundamentales extraídas período a período. Sin embargo, el rango de extracción del tono, esta definida por cualquier búsqueda para períodos

de 70 - 625 Hz o 200 – 1000 Hz. Por lo tanto, el rango alto lo determinará una fundamental bajo los 200 Hz. El cálculo se realiza mediante la siguiente fórmula:

$$Flo = \min\{Fo^{(i)}\}, i = 1, 2, \dots, N \quad Fo^{(i)} = \frac{1}{To^{(i)}} \quad (13)$$

Donde: $To^{(i)}$ [$i = 1, 2, \dots, N$] = Período de Fo ; información del período del tono extraído
 N = Número de períodos del tono extraído.

Fhi: Frecuencia Fundamental Mayor [Hz] - Es la máxima Fo obtenida a partir de la información del tono extraído período a período. Sin embargo, el rango de extracción del tono, está definida por cualquier búsqueda para períodos de 70 – 625 Hz o 200 – 1000 Hz. Por lo tanto, el rango normal lo determinará una fundamental sobre los 625Hz. Su cálculo es:

$$Fhi = \max\{Fo^{(i)}\}, i = 1, 2, \dots, N \quad Fo^{(i)} = \frac{1}{To^{(i)}} \quad (14)$$

Donde: $To^{(i)}$ [$i = 1, 2, \dots, N$] = Período de Fo ; información del período del tono extraído
 N = Número de períodos del tono extraído.

STD: Desviación estándar de la frecuencia fundamental [Hz] - Desviación estándar de todos los valores extraídos de la muestra de voz período a período. Es calculado como la desviación estándar de la información de Fo :

$$STD = S = \sqrt{\frac{1}{N} \sum_{i=1}^N (Fo - Fo^{(i)})^2} \quad (15)$$

Donde: N = numeración de periodos del tono extraído.

To: Período tonal promedio [mseg] - El valor promedio de todos los valores del período del tono extraído. Se calcula por medio de:

$$T_o = \frac{1}{N} \sum_{i=1}^N T_o^{(i)} \quad (16)$$

Donde: $T_o^{(i)}$ [$i = 1, 2, \dots, N$] = Período de F_o ; información del período del tono extraído
 N = número de períodos del tono extraído.

4.3.1.4 Ruido y Armónicos

NHR: Razón Ruido-Armónicos – Relación entre la energía espectral inarmónica en el rango de frecuencia de 1500-4500Hz y la energía espectral armónica en el rango de frecuencia de 70-4500Hz. Ésta es una evaluación general del ruido presente en la señal analizada. Un aumento en los valores del NHR se interpreta como un aumento en el ruido espectral el cual puede ser debido a variaciones en la amplitud y frecuencia, ruido de turbulencia, componentes subarmónicos y/o quiebres en la voz. El grado de periodicidad acústica (razón ruido-armónico) se calcula o está expresada en dB. Si el 99 % de la energía de una señal es una parte periódica, y el 1 % es ruido, el NHR se expresa como:

$$20 \text{ (dB)} = 10 * \log (99/1) \quad (17)$$

Por lo tanto, se admite que 20 dB corresponde al umbral de normalidad. Si tenemos entonces un NHR de 0 dB representa igualdad de energía en los armónicos y en el ruido de la señal, es decir, asumimos una patología en el hablante.

DUV: Grado de pérdida de voz [%] – Se refiere a una evaluación estimada relativa de áreas no-armónicas (donde F_0 no puede ser detectada) en una muestra de voz. Es calculada como una razón entre el número de segmentos auto-correlacionados donde una decisión de pérdida de voz fue hecha y el número total de segmentos auto-correlacionados. Mide la habilidad de la voz para sostener expresiones continuas. El umbral normado es 0, porque una voz normal, en la definición de expresión sostenida, cualquier segmento no debe tener pérdida de voz.

DSH: Grado de componentes sub-armónicos [%] – Es la evaluación estimada de sub-armónicos para componentes de F_0 en una muestra de voz. Es calculada como una razón entre el número de segmentos auto-correlacionados donde el tono hallado es un sub-armónico del tono real y el número total de segmentos auto-correlacionados. El grado de componentes sub-armónicos en voces normales debe ser igual a 0. Esto es cada vez más esperado para voces donde períodos dobles o triples reemplazan a la fundamental en ciertos segmentos fuera de la longitud de análisis. Estos efectos son típicos para voces diplofónicas y voces con estrechez glotal.

4.3.1.5 Quiebre de la voz

DVB: Grado de quiebre de la voz [%] - Razón entre la longitud total de áreas (en tiempo) que representan quiebres en la voz y el tiempo de una muestra de voz completa. Mide la habilidad de la voz para sostener expresiones continuas. El umbral normado es 0, porque una voz normal, durante el trabajo de sostener la voz, no debe tener áreas de quiebre de la voz. Es calculado como la razón entre la suma de todas las longitudes de quiebres de voz (T_1, T_2, \dots) y la longitud de la muestra completa (T_{sam}).

$$DVB = \frac{t_1 + t_2 + \dots + t_n}{T_{sam}} \quad (18)$$

NVB: Número de quiebres de la voz - Número de veces en que el período de la fundamental es interrumpida durante la muestra de voz (medido desde el 1º período detectado hasta el último). Mide la capacidad de la voz para sostener una expresión hablada sin interrupciones.

En base a lo anteriormente expuesto, podemos señalar que existen un sin número de parámetros o descriptores capaces de entregarnos en detalle el comportamiento de una voz.

4.3.2 Tipos de Patologías

- ✓ **Alteraciones de la Resonancia:** El sonido glótico es débil y monótono y desde el punto de vista fisiológico debe atravesar los espacios resonadores para obtener riqueza en las formantes. Este sonido glótico, al pasar por los resonadores concentra energía en diversas frecuencias (formantes) y extiende los valores de la frecuencia fundamental. En el caso de resonancia agudas la laringe asciende y para resonancias graves la laringe desciende.
 - § **Hiponasalidad:** o Rinolalia abierta, es cuando el paladar blando no cierra y todos los sonidos tienen resonancia nasal, no sólo la m, la n o la ñ. Para comprobar ésta alteración se hacen pruebas específicas emitiendo frases con consonantes nasales.
 - § **Hipernasalidad:** es cuando el paladar está prácticamente cerrado y ni siquiera las consonantes nasales tienen resonancia nasal. Se comprueba de la misma manera que la Hiponasalidad.

- ✓ **Fatiga vocal:** Es la incapacidad para fonar durante períodos largos de tiempo sin cambiar el timbre vocal, es decir en la fatiga vocal hay molestia o dolor laríngea acompañada de cambios de timbre tono vocal de forma que la voz se percibe más grave.

- ✓ **Dureza de la voz:** También llamada tensión, corresponde al fenómeno de hiperfonación excesiva en la laringe. Este se debe a un exceso en la función laríngea, que en algunos casos se les pueden considerar como lesiones funcionales a los nódulos laríngeos.

- ✓ **Aire en la voz:** Se refiere a la sensación de voz aérea o voz soplada, y es producida por escape de aire de las cuerdas vocales, sea cual fuese la causa que las determina, es decir, si por alguna razón no existe cierre glótico parte del aire espirado se pierde de forma turbulenta entre las cuerdas.

- ✓ **Disfonía:** Es una voz anormal, percibida como tal por el oyente y que puede afectar al tono, la intensidad, el timbre, la fluidez o una combinación de estos elementos. En otras palabras la presencia de vibraciones irregulares e incluso la ausencia parcial de ellas juntos con los defectos el cierre glótico significan un defecto de funcionamiento.

- ✓ **Voz Ronca:** También llamada disfonía por antonomasia, corresponde al defecto de vibración en las cuerdas vocales parcial o total. La forma más simple de ronquera es el bloqueo instantáneo de la vibración por la interposición de una mucosidad sobre el borde libre, y la forma más grave es la infiltración tumoral de la glotis.

- ✓ **Alteraciones en la flexibilidad:** En una voz normal durante el habla contextualizada posee unas variaciones en la frecuencia fundamental, intensidad y el timbre mediante las que se expresan significados emocionales e intelectuales. En las voces patológicas esas fluctuaciones están disminuidas y aplanadas, o son literalmente excesiva.[6]

4.4 Aplicaciones Lingüísticas

4.4.1 Orígenes del español

El castellano, dialecto románico surgido en Castilla y origen de la lengua española, nació en una franja montañosa, mal y tardíamente romanizada, inculta y con fuertes raíces prerromanas (Burgos, Iria Flavia, Oviedo, Amaya, Pamplona), en la cual surgieron los condados y reinos medievales españoles, y en torno a esos nuevos centros fueron desarrollándose las variedades dialectales. El castellano, dialecto de los montañeses y vascos encargados, en el siglo IX, de defender de los árabes (en la península desde el año 711) la frontera oriental del reino asturleonés, toma su nombre de castilla que en periodo visigótico significó pequeño campamento militar (diminutivo de castrum) y luego tierra de castillos. Con respecto a los vascos, se sostiene que éstos, con su propia lengua, influyó profundamente en esta nueva lengua románica.

La modalidad idiomática navarro-aragonesa, utilizada en el lugar en donde confluían tres reinos, Castilla, Navarra y Aragón, dio origen, en el siglo XI, a los primeros documentos peninsulares en una lengua romance: las glosas emilianenses. En el año 1042, por otra parte, se escribieron las jarchas, primeros textos en castellano, pero con caracteres árabes o hebreos.[21]

4.4.2 El español llega a América

En 1492, cuando Cristóbal Colón llegó a América, el castellano se encontraba consolidado en la península, pero durante los siglos XV y XVI se produjo una verdadera revolución consonántica que afectó especialmente a las llamadas sibilantes, las cuales se redujeron, y ésa fue la variedad que llegó al Nuevo Mundo, generalmente conocida como español de América. En éste continente se enriqueció con el aporte de las lenguas aborígenes de Hispanoamérica.

A partir del siglo XVI se impuso el término de español al convertirse en lengua nacional. De hecho, en 1536, es el emperador Carlos I, en presencia del Papa, quien utiliza por primera vez la expresión lengua española, la cual —según el monarca— "era tan noble que merecía ser sabida y entendida de toda la gente cristiana, hecha para hablar con Dios". El término castellano subsistió como nombre del actual dialecto de Castilla.

Desde 1492 a la fecha, el español se ha extendido por los cinco continentes. Además de ser la lengua oficial de España y de diecinueve países de América y el Caribe (México, Guatemala, Honduras, El Salvador, Cuba, República Dominicana, Nicaragua, Costa Rica, Panamá, Venezuela, Colombia, Ecuador, Perú, Bolivia, Paraguay, Uruguay, Argentina, Chile y Puerto Rico, en este último, junto con el inglés), es, además, idioma nativo —total o parcialmente— en determinadas zonas de Estados Unidos (Nuevo México, Arizona, Texas, California y Florida), país en el cual es la segunda lengua principal, con 23 millones de hablantes. Se estima que en la próxima década, entre 27 y 30 millones de norteamericanos hablarán español, los cuales constituirán el 12% de la población de Estados Unidos.[22]

4.4.3 Zonas lingüísticas americanas

El sistema educacional fue, quizás, uno de los factores determinantes en el establecimiento de diferencias lingüísticas, pues ya en 1538 la escuela de Santo Domingo se convirtió en la Universidad de Santo Tomás de Aquino, y en la misma ciudad se creó la Universidad de Santiago de la Paz en 1540, mientras que la Universidad de Córdoba (Argentina) fue creada en 1613. Ahora bien, otra de las causas de la diferenciación dialectal se refiere a la época de la colonización, ya que la ciudad más antigua, Santo Domingo, fue fundada casi en el momento de la llegada de Colón a América, mientras que Montevideo se fundó en 1722.

Sin embargo, estos intentos de zonificación no siempre han sido fecundos, debido a que no se cuenta con datos precisos o suficientes en cada lugar, por

ejemplo, a través de la elaboración de atlas lingüísticos; pero, a pesar de lo anterior, algunos autores coinciden en distinguir las siguientes zonas: 1) México y sur de los Estados Unidos, 2) Caribe, 3) zona andina, 4) zona rioplatense y 5) zona chilena. Éste proyecto de investigación, la cual estamos centrados, comprende únicamente a la zona chilena, y sus respectivas subdivisiones, es decir, zona norte, zona centro y zona sur.

Considerando los elementos prosódicos o los elementos suprasegmentales (acento, entonación, melodía, ritmo, etc.) se puede decir que han recibido una mayor atención en los últimos años, no únicamente por su interés teórico y descriptivo, sino también por sus implicaciones en lo que respecta a mejoras y naturalidad de sistemas de conversión de texto a habla que pueden obtenerse mediante un modelamiento prosódico lo más cercano posible al habla natural.

Estos elementos suprasegmentales, lejos de constituir un adorno, son precisamente los que organizan el hilo de sonido que percibimos. Si el sonido no estuviera agrupado de manera significativa, no habría comunicación lingüística; se oirían gritos o murmullos. El oyente recibe, cuando se comunica lingüísticamente, segmentos sonoros relativos a las unidades de información que le envía el hablante. Con el sonido, va una serie de informaciones de naturaleza prosódica, no solamente con información referencial sino también con información dialectal, sociolingüística e inclusive, emotiva.

Al percibir el habla, tenemos una melodía conformada por las variaciones de frecuencia de la fundamental, pero al mismo tiempo sentimos ritmo, pausas, intensidad y otros elementos fónicos para algunos autores, la prosodia o entonación se componen de todos estos elementos. Para otros, la entonación está conformada por todos esos elementos que "suenan" con la fundamental y es sólo uno de los elementos de la prosodia. León (1996) y Mora (1996) entre otros denominan entonación al movimiento melódico, es decir, a las variaciones de la frecuencia fundamental F_0 . La prosodia estudia, además de la entonación, el acento y el ritmo, que incluye pausas, timbre y velocidad de habla.

El análisis y descripción del habla continua (habla espontánea) no es una línea de investigación nueva dentro de la Fonética, pues hace ya algunos años se vienen publicando trabajos que estudian aspectos tales como la influencia de la velocidad de elocución sobre la duración de los sonidos, la naturaleza y función de las pausas en el habla, lo que se podría denominar las unidades de procesamiento de la información, etc. Lo que sí supone una novedad y un aporte, es el estudio sistemático de los fenómenos que se producen en el habla espontánea y su reducción a una serie de reglas que permitan predecir la variabilidad. Estos fenómenos, como dijimos anteriormente básicamente son el **Acento, la Entonación y la Melodía**¹. [7]

4.4.4 Fenómenos producidos en el habla espontánea

4.4.4.1 Acento

Rasgo prosódico que permite poner en relieve una unidad lingüística superior al fonema (silaba, palabra, frase) para diferenciarla de otras unidades lingüísticas del mismo nivel. El acento se manifiesta como un contraste entre unidades acentuadas y unidades inacentuadas, pero acústicamente lo hace por medio de la melodía.

4.4.4.2 Melodía

Variación de la frecuencia de vibración de las cuerdas vocales, percibidas como variaciones frecuenciales (Pitch) a lo largo de un enunciado. La melodía es representada acústicamente por las curvas melódicas.

¹ Vale decir, que ésta es una visión del autor, y no concuerdan con la clasificación que realizan otros autores.

4.4.4.3 Entonación

La entonación se puede definir como variaciones frecuenciales a lo largo de un enunciado, que es el resultado de la integración de la melodía y el acento.

4.4.5 Detección de F_0

En el estudio de la entonación, el principal problema es el estudio de las curvas melódicas. Una de las primeras dificultades para el análisis prosódico del habla es el proceso de detección de la frecuencia fundamental.[5]

Existen diferentes técnicas para la obtención de la frecuencia fundamental, que pueden agruparse en dos grandes tipos: las que actúan en el dominio temporal y la que lo hacen en el dominio frecuencial. Las que actúan en el dominio temporal intentan reconocer en la onda sonora formas periódicas recurrentes. Por otro lado los procedimientos basados en el análisis frecuencial no actúan directamente sobre la forma de onda, si no que determinan la distancia existente entre los armónicos del espectro. Algunos ejemplos de este tipo son la autocorrelación, la compresión espectral o la comparación de los armónicos.

Las curvas melódicas que se obtienen tanto con los sistemas frecuenciales como temporales plantean errores, que son necesarios eliminarlos o minimizarlos. A continuación se describirán algunos procedimientos para optimizar los resultados.

4.4.5.1 Minimización de los errores de detección

Las técnicas más comunes para la minimización de errores son la restricción del rango de búsqueda y el alisado:

- ✓ **Rango de búsqueda:** una forma de eliminar los errores de detección, especialmente los ocasionados por la duplicación y omisión de periodos, es delimitar el rango frecuencial de búsqueda

en función del rango frecuencial de la muestra a analizar para descartar directamente los valores situados por encima de o por debajo del umbral mínimo y máximo.

Hombres = 80 Hz – 300 Hz

Mujeres = 130 Hz – 525 Hz

▼ **Alisado:** ésta técnica consta de dos fases:

§ Se compara cada valor de F_0 con los valores anteriores y/o posteriores y se establece un porcentaje máximo de variación con respecto a los valores adyacentes, considerando así que aquellos valores de F_0 que superan el porcentaje de variación, constituyen errores de estimación. Tanto el número como el porcentaje de variación son específicos para cada programa.

§ Los valores de F_0 considerados como errores de detección se aproximan a los valores adyacentes.

4.4.6 Estilización

Los detectores de la frecuencia fundamental cometen errores que dificultan las interpretaciones de las curvas melódicas y que es necesario eliminar. En segundo lugar las curvas melódicas contienen variaciones o rupturas relacionadas con las características intrínsecas de los elementos segmentales que componen los enunciados, como son las interrupciones en la curva debido a la presencia de segmentos sordos, o las pequeñas variaciones en el curso de la curva melódica debidas a la aparición de determinados alófonos.[4]

Estas variaciones llamadas micro melodías, no aportan información lingüística relacionada con la interpretación de la curva melódica, aunque pueden

constituir un indicio adicional para el reconocimiento del segmento en cuestión. Pueden ser de dos formas.[5]

✓ **Interrupciones debidas a la aparición de segmentos sordos a lo largo del enunciado.**

Los segmentos sordos al carecer de F_0 , provocan interrupciones en las curvas melódicas que pueden causar problemas en el proceso de interpretación. Estas interrupciones no son perceptibles para el oyente, que posiblemente realiza el proceso de reconstrucción del contorno, y no son pertinentes en el estudio de la entonación desde el punto de vista de la lingüística.

✓ **Variaciones de F_0 debidas a la naturaleza de los elementos segmentales.**

Estas variaciones, al igual que las relacionadas con la sonoridad, afectan a la forma de la curva melódica, y bajo éste punto de vista (naturaleza de los elementos segmentales), encontramos tres criterios:

- § El llamado fundamental intrínseco de las vocales, que depende del grado de abertura de las mismas, de forma que cuanto más cerrada es la vocal, más alta es la frecuencia fundamental.
- § Los pequeños descensos en la frecuencia fundamental ocasionados por las características de cierta clase de consonantes, como por ejemplo las vibrantes y las aproximantes.
- § Los ascensos y descensos que se observan antes y después de un segmento sordo, debido a los efectos de coarticulación.

La estilización y la codificación son en general los procedimientos utilizados para suplir éstas anomalías. La codificación implica normalmente una aproximación más fonológica al estudio de la entonación, mientras que la estilización es más fonética.

En la estilización, las curvas melódicas quedan reducidas a una serie discreta de puntos, que unidos mediante líneas, proporcionan una representación de sus movimientos más relevantes.

Según el número de puntos que se utilice para la representación será ancha (detección de un número menor de puntos) o estrecha (detección de un número mayor de puntos). En general una estilización basada en criterios de tipo lingüísticos, en la que se obtiene un valor en la frecuencia fundamental para cada sílaba, por ejemplo, será más estrecha que una estilización basada en criterios acústicos, donde se detecta un valor de F_0 , denominado **punto de inflexión**, al observar un cambio importante en la dirección y la pendiente de la curva. A continuación describiremos algunas maneras de estilización.

4.4.6.1 Estilización manual

Se basa en las diferencias frecuenciales entre los diferentes puntos de la curva melódica. Dado que los segmentos vocálicos son los que se ven menos afectados por las variaciones micromelódicas, los valores se pueden tomar en algunos puntos de la curva melódica coincidentes con las vocales del enunciado. Sin embargo hay variaciones de F_0 que se producen dentro de una misma vocal que con éste método puede que no queden reflejadas en la estilización resultante.

4.4.6.2 Estilización automática

Al igual que el método manual convierte la curva en una serie de puntos de inflexión, unidos mediante líneas, este sistema determina los puntos de inflexión por medio de un cálculo de regresión que aproxima la forma de la curva de F_0 y elimina las variaciones irrelevantes.

4.4.7 Análisis de patrones melódicos

Por patrón melódico se entiende una representación estilizada de las curvas melódicas que recoge las características comunes a un conjunto de estilizaciones que comparten determinadas propiedades fonéticas y/o lingüísticas. Los patrones melódicos pueden definirse en dos grandes tipos: los patrones locales que serían patrones de ámbitos reducidos (uno o dos grupos acentúales como máximo) y los patrones globales, de ámbito más extenso (grupos melódicos, frases o párrafos), a los cuales se superpondrían los patrones locales para componer la curva melódica final.

4.4.7.1 Patrones locales

Una forma de representar los patrones locales a partir de contornos estilizados, es en forma de series de puntos de inflexión unidos por movimientos ascendente, descendentes y planos. Estas representaciones son válidas tanto para movimientos en posiciones finales (patrones finales) como movimientos en posiciones iniciales (patrones no finales).

4.4.7.1.1 Patrones no Finales

La forma típica de los movimientos no finales consiste en una primera fase de descenso de la F_0 , seguida de una fase de ascenso. El punto de inflexión que

marca el fin de la fase de descenso y el inicio de la fase de ascenso suele situarse al principio de la sílaba tónica del grupo acentual.

4.4.7.1.2 Patrones Finales

Las descripciones tradicionales de la entonación del español se centran en la forma de los movimientos finales. Sin embargo, en base a patrones melódicos finales establecidos se pueden agrupar en cuatro subcategorías: descendente, ascendente, ascendente-descendente y descendente-ascendente.

4.4.7.2 Patrones globales

La definición anteriormente realizada de la entonación relaciona ésta con la forma global de una curva melódica a lo largo de un grupo fónico, es decir con su posición global. Los patrones para la frase corta simple en español se basan en ésta idea.

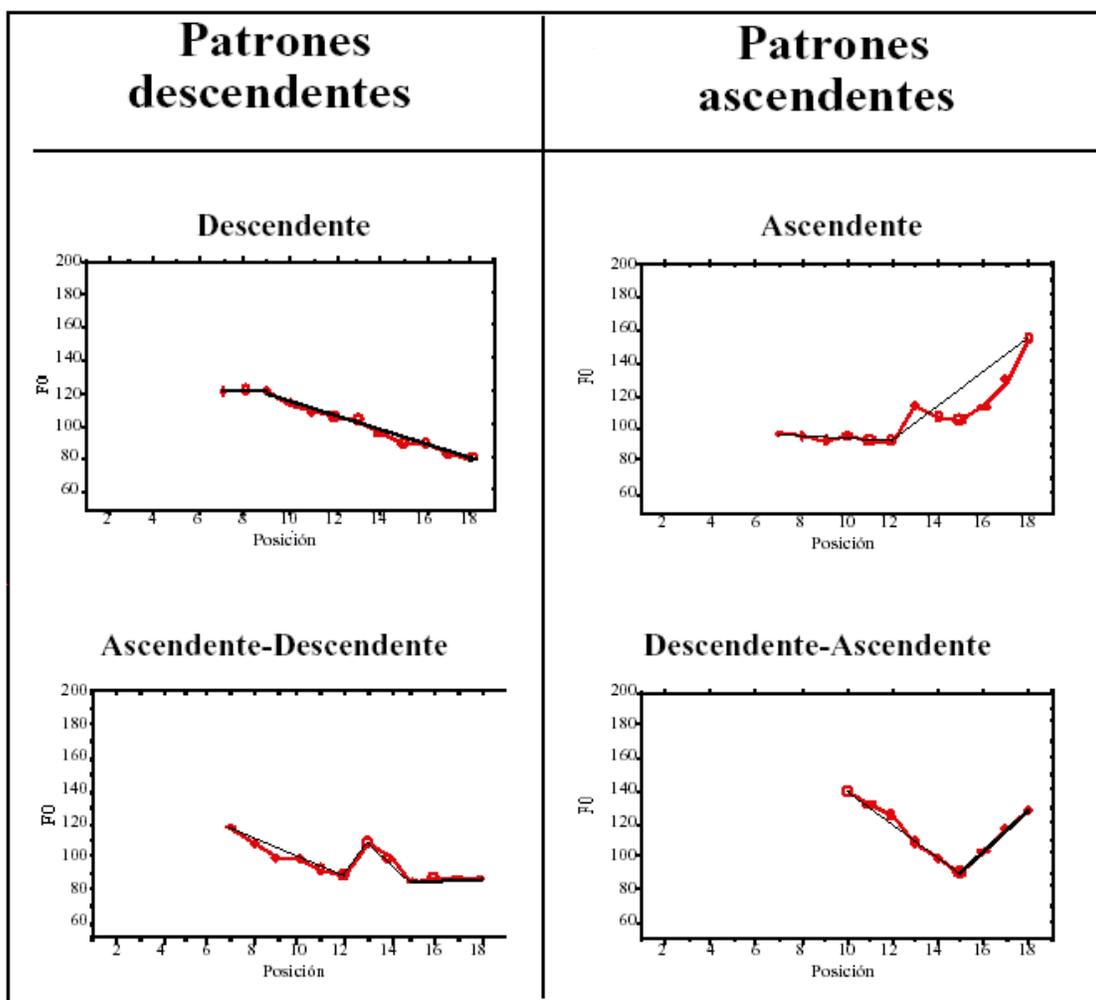


Fig. 4.17 – Clasificación de los patrones locales para el español. (Fuente: [8]).

5 DESARROLLO EXPERIMENTAL

5.1 Implementación de sistema electroacústico

La creación de una cadena electroacústica permitirá obtener las grabaciones que formarán parte de la Base de Datos que será construida, buscando que sea lo más óptima posible para obtener un buen registro de las señales y a un bajo costo para las personas que participan de éste proyecto (tesistas).

Dicha cadena está conformada con equipamiento habilitado tanto por nuestra casa de estudios como también por la Estación Televisiva Chilevisión. El esquema de la cadena Electroacústica utilizado se muestra a continuación:

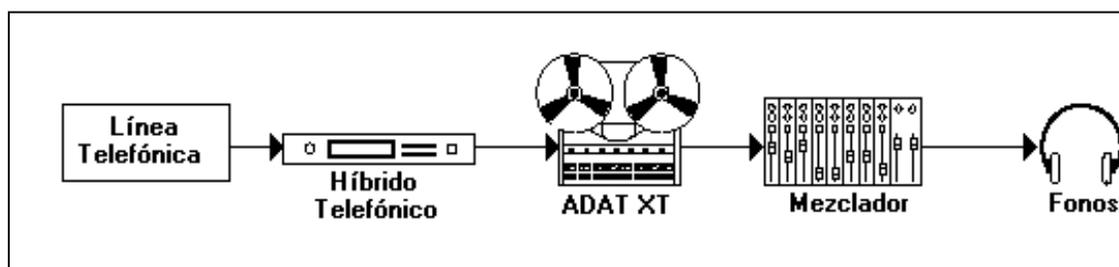


Fig. 5.1 – Sistema Electroacústico Grabación. (Fuente: Elaboración propia).

La señal que se desea grabar es la proveniente de la línea telefónica, donde será necesario utilizar para ello una interfaz telefónica. La interfaz utilizada fue la **AC-10H Clear-Com**. Ésta mediante cambios de impedancia, nos permitirá interconectar la línea telefónica con nuestra cadena de grabación; además de entregarnos la mejor calidad de sonido de la línea telefónica.

La señal de audio obtenida por la interfaz, pasa a la cadena de grabación por medio de un jack estéreo a la entrada balanceada del ADAT Alesis XT20, donde quedará registrada en un video cassette S-VHS con su respectivo formato digital. El mezclador Shure SCM810, por medio de fonos, nos permitirá la función de monitorear lo que está siendo registrado en el ADAT Alesis XT20, conectando la salida del canal que se está grabando a la entrada de línea de uno de los ocho canales que éste mezclador posee, pudiendo variar a gusto el nivel de monitoreo de las señal. Este monitoreo se hace con el fin de corroborar si la información registrada es acorde y cumple con las necesidades para el desarrollo de nuestro trabajo.

Finalmente, el sistema quedó habilitado en una dependencia de nuestra casa de estudios para poder realizar las grabaciones y los contactos de las personas que formarán parte de la Base de Datos.

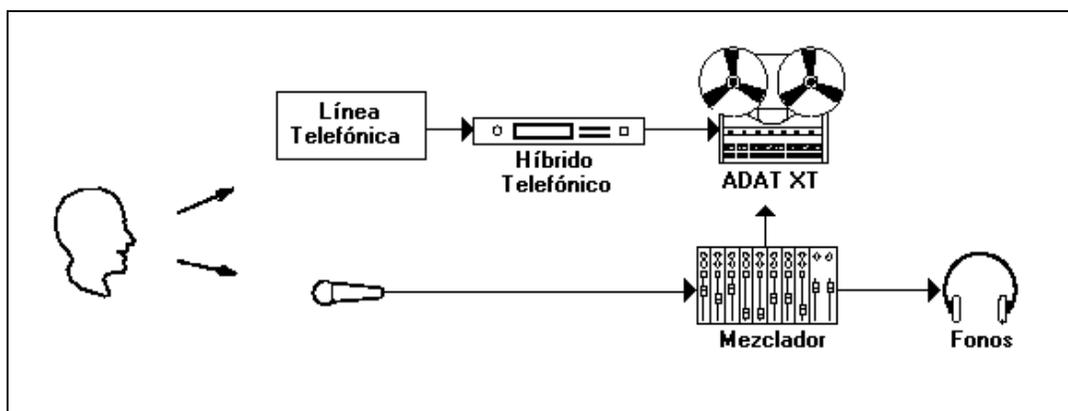
5.1.1 Implementación de sistema electroacústico secundario

Fue necesario incorporar una segunda cadena electroacústica con el objetivo de determinar como actúa la interfaz telefónica sobre la voz humana.

Mediante una comparación entre las señales grabadas directamente por micrófono y las señales grabadas vía telefónica, se encuentra la respuesta de frecuencia de la interfaz telefónica para dichas señales, determinando la variación existente entre la información registrada por el medio telefónico con respecto a un medio de grabación que entrega un registro totalmente fiable como lo es el micrófono.

La comparación se lleva a cabo en ambiente computacional extrayendo los parámetros de la voz que son de interés de las muestras registradas.

Para esto se realizaron grabaciones con un micrófono direccional a condensador Shure KSM141 a cuatro personas con el siguiente sistema en adición al sistema electroacústico principal:



*Fig. 5.2 – Sistema electroacústico secundario de grabación simultáneas de micrófono y teléfono.
(Fuente: Elaboración propia).*

Se puede notar en la figura 5.2 que la persona al momento de hablar queda simultáneamente registrada por medio de la interfaz telefónica como también por el micrófono Shure KSM141. Ambos registros se almacenan en pistas adyacentes del video cassette S-VHS por medio del ADAT Alesis XT20, obteniendo la misma información pero registrada de distintos medios.

Una vez que las señales de ambas pistas se traspasan de formato digital S-VHS a un ambiente computacional, se obtienen dos archivos WAV por persona que corresponden al registro de micrófono y al telefónico.

Por medio de programas adecuados que tienen relación con el análisis de la voz, son extraídos separadamente, a éstos archivos, algunos parámetros acústicos, tales como frecuencia fundamental, formantes, algunos descriptores acústicos como Shimmer, Jitter y NHR. Los resultados que se obtienen para llevar a cabo esta comparación se muestran más adelante dentro de la sección **5.6 Edición** en el ítem **5.6.3 Resultados comparativos de mediciones micrófono- teléfono**, donde se entregan las variaciones a que están sometidos los distintos parámetros que se evaluarán, a causa de los distintos medios de registro.

5.2 Planificación de las Grabaciones para la Base de Datos

5.2.1 Marcha Blanca

Antes de comenzar con las grabaciones que serán parte fundamental en nuestra Base de Datos, fue fijado previamente un período de pruebas del sistema electroacústico implementado, éste se denomina *Marcha Blanca*, la cual busca prever cualquier posible problema que pueda surgir en el transcurso de las grabaciones de la Base de Datos.

En el transcurso de la Marcha Blanca, nos se familiarizamos con el funcionamiento del equipamiento y se optimizó la metodología de trabajo; esto se hizo simulando la situación real de grabación de personas; para lo cual se grabaron personas sólo de Santiago anteriormente contactadas y que poseían el instructivo.

De ésta manera fue posible estructurar la forma de trabajo con la que se realizarían las grabaciones para la Base de Datos, tomando en cuenta factores como el modo de grabación y el tiempo de contacto con las personas a grabar, la calendarización del contacto con las mismas, la manera de hacerles llegar el instructivo y su adecuada explicación para que al momento de ser grabada la persona cometa el mínimo de errores.

Otro factor que se tomó en cuenta es el tiempo de grabación promedio de cada persona registrada, para calcular cuanto tiempo se utilizará por locutor en la pista de la cinta S-VHS que esté siendo utilizada al momento de grabar, buscando optimizar los tiempos de cada llamado telefónico en función de lograr una optimización de los costos.

5.2.2 Número de personas que formarán parte de la Base de Datos

Para el desarrollo de la base de datos, se determinó registrar a un total de 300 personas a lo largo de todo Chile, realizando una distribución uniforme sobre el número de registros que se harían por región. Se grabaron, 60 personas en la Región Metropolitana y 20 personas a lo largo de cada región del país, entre hombres y mujeres, donde estos dos géneros no están distribuidos de igual forma.

Los registros hechos son totalmente al azar, o sea no se busca registrar personas específicas en cuanto a su tipo de voz, sino que lograr tener un registro lo más heterogéneo y representativo de la población chilena .

5.2.3 Requisitos de la persona

La persona que acepte formar parte de la Base de Datos lo debe hacer bajo libre elección y en forma anónima, debido a que la Base de Datos no busca identificar por el nombre a cada persona que esté registrada en ella, pero si hacer una discriminación de acuerdo al lugar donde habita, los estudios y la actividad que realiza la persona.

Además la persona que será grabada, debe cumplir una serie de requisitos que se listan a continuación:

- ✓ Tener entre 20 y 51 años de edad.
Se tomo éste rango de edad por que los especialistas consideran que la voz desde un punto de vista fisiológico está formada.

- ✓ Tener cursada como mínimo la Enseñanza Básica.
Ésta consideración se relaciona con el simple hecho de que la persona pueda leer el instructivo que le será enviado.

- ✓ Tener residencia por lo menos desde hace 10 años en el lugar al que dice pertenecer.

Finalmente éste requisito se asocia a la idea de considerar personas cuya forma de hablar y de expresarse sea la correspondiente a la zona.

5.2.4 Contactos

Para contactar a lo largo de todo Chile a las personas que posiblemente formarán parte de la base de datos, se comenzó contactando vía telefónica a los departamentos de asistencia social pertenecientes a las municipalidades regionales del país, con el objeto de encontrar voluntarios que quisieran formar parte de esta Base de Datos. Otros contactos fueron hechos por medio de los mismos tesistas, los cuales utilizaron contactos con familiares y conocidos.

A todos los contactos se les informó y explicó a gran escala en algunos casos y detallados en otros, en que consistía el trabajo, la institución que lo avalaba, hacia que área estaba enfocado, las utilidades que podría traer su desarrollo a estas áreas de interés, además de dejar en claro que su posible participación sería de forma anónima.

Vale reiterar que todos los contactos hechos debían cumplir con los requisitos previamente dados, bajo un criterio de diversidad de selección de las posibles personas que accedieran a ser registradas; o sea no centralizado sobre un estrato social o sector de población específico del país.

5.3 Instructivo

El instructivo es la pauta que deberán seguir las personas contactadas al momento de ser grabadas, previa explicación del encargado de registrar las llamadas, buscando que la información que sea registrada por persona, sea lo más óptima y útil posible para el desarrollo del trabajo. La información registrada será utilizada para distintos tipos de estudios, donde se podrán evaluar a las personas registradas bajo la misma referencia, que en este caso es el instructivo.

La creación y estructuración del instructivo, fueron delegadas sobre un equipo de Fonoaudiólogos especialistas en voz normal, patologías del habla y lingüística pertenecientes a la Universidad Mayor.

El instructivo consta de 5 partes, en la primera el locutor debe decir la ciudad y zona en que habita, el nivel de estudios que posee y actividad que realiza, lo que servirá para ordenar e identificar a las personas que estén en la Base de Datos. Las siguientes cuatro partes del instructivo que el locutor debe seguir al momento de ser grabado, representan las áreas de estudios que se pueden desarrollar basándose en los registros de la Base de Datos. Estas cuatro partes son:

- ✓ Vocales
- ✓ Frases
- ✓ Números
- ✓ Prosodia

5.3.1 Vocales

Al momento de ser grabado, el locutor debe decir las 5 vocales en forma sostenida manteniéndolas un par de segundos, de forma no solapada y manteniendo un mismo tono durante la ejecución de éstas.

La información registrada será de utilidad para realizar estudios sobre la emisión normal o patológica de la voz, como también para estudios de la articulación vocálica.

En el estudio de la emisión vocal normal o patológica, de las personas pertenecientes a la Base de datos, se evaluarán parámetros sonoros de la voz, mediante el análisis de sonidos fonados como lo son las vocales, es decir, con vibraciones en las cuerdas vocales. Algunos parámetros que se analizarán para esos estudios son la frecuencia fundamental, además de parámetros especializados en voz patológica como son el NHR, Shimmer y Jitter.

Para el estudio la de articulación vocálica, a partir de estos sonidos fonados, se analizará la carta de Formantes, el cual es un gráfico que entrega la relación existente entre el primer y segundo formante, esto permite diferenciar la articulación asociada a las distintas vocales ya sea de una persona como de toda la población registrada en la Base de Datos.

5.3.2 Frases

A continuación de las Vocales el locutor debe leer las siguientes dos frases:

- 1) ***“Las personas que habitan la región Metropolitana, padecen frecuentemente de enfermedades respiratorias”***
- 2) ***“Se dice que los niños gozan de mayor y menor manera los obsequios entregados sin razón alguna”***

La lectura de estas dos frases debe ser en forma natural y pausada, buscando que cada palabra de la frase sea dicha de la mejor forma posible. La información recabada permitirá realizar estudios sobre los fonemas del habla española chilena además de la coarticulación que exista entre ellos.

Estos análisis se harán sobre la base de las dos frases anteriormente expuestas, las cuales poseen información fonéticamente balanceada; es decir, que contienen todos los fonemas del español de Chile, los que serán analizados de acuerdo a la estructura formántica de cada uno con la ayuda de programas computacionales. No está demás señalar que una persona al leer una serie automática ajusta su voz a un rango más cómodo, de manera tal que permita una fácil lectura.

5.3.3 Números

Seguido de la lectura de las frases, el locutor debe decir los números del 1 al 10 de forma natural y seguida dejando una pequeña pausa entre éstas de manera que no se unan.

Esta serie automática ejecutada, se utiliza para realizar estudios de la altura tonal, o tono medio hablado, obteniendo un valor promedio de la frecuencia fundamental asociada a la vibración de las cuerdas vocales de cada persona. Esto sucede por que en las series automáticas las personas ajustan su voz al rango de frecuencia que más le acomode.

Existen valores que se consideran normales para hombres y mujeres obtenidos de estudios especializados en esta área, los cuales serán comparados con los registros de la Base de Datos que busca estandarizar el tono medio hablado del hablante chileno.

5.3.4 Prosodia

Como último paso dentro la grabación, el locutor debe efectuar un breve comentario de la forma que celebran en la ciudad que habita, las festividades del 18 de Septiembre –Aniversario de la Independencia de Chile-, destinando un tiempo máximo de 15 segundos para dicho relato. Este comentario debe ser efectuado de la forma más espontánea posible por la persona, poniéndose en la situación de estar conversando vía telefónica con un familiar o alguien conocido.

Esta información permite el estudio de la prosodia de las personas registradas en la Base de Datos, evaluando las variaciones que existen en el tiempo de la frecuencia fundamental sobre el relato de habla espontánea ejecutado por cada locutor. Se buscará determinar la entonación de la voz por medio de las frases, y su acentuación por medio de las palabras, lo que entrega curvas melódicas asociadas a las personas con el fin de encontrar las diferencias fonéticas propias del lenguaje para distintas zonas de nuestro país.

5.4 Grabación de Señales

Todos los registros de la Base de Datos fueron grabados en videocasetes de formato digital S-VHS, lo cual se realizó por medio del ADAT Alesis XT20. Dichos registros, se almacenaron con una frecuencia de muestreo de 48 Khz., con una cuantización de 16 bits por muestra y en formato monofónico.

Los antecedentes entregados por la Marcha Blanca permitieron desarrollar un método de grabación ordenado, rápido y preciso que fue aplicado a lo largo de todo el proceso de recopilación de señales para la Base de Datos.

Al tiempo en que cada persona era contactada, se fija de mutuo acuerdo la hora y fecha en que se realizará la grabación, así como también la forma en que se enviará el instructivo, el cual generalmente se les hizo llegar vía Fax o vía correo electrónico. Con esta calendarización global, se crearon horarios de grabación los cuales fueron distribuidos entre los alumnos tesistas, según la disponibilidad de éstos para realizar las grabaciones.

En el momento de la grabación, se procuró que las señales de audio fuesen lo más fielmente registradas, tomando en cuenta la previa explicación del instructivo a seguir y el sistema de grabación que se tiene para este fin. De este proceso se obtuvieron un total de 2 videocasetes de S-VHS; cada uno con 8 pistas de 40 minutos de grabación registrándose finalmente un total de 12 pistas con la información de todos los locutores en forma consecutiva.

La forma de identificar el lugar de la cinta y la pista en que se encontraba cada locutor registrado, se hizo gracias a un registro en planilla Excel que se iba completando conforme se realizaban las grabaciones, esta planilla Excel es de gran utilidad a futuro cuando se comienza la edición de todas las muestras ya que posee la información de cada grabación realizada, como el número con que se identifica al locutor, ciudad donde habita, estudios, actividad que realiza, sexo, tiempo de inicio y término de la grabación, pista en que se encuentra registrado así como también el número de cinta. La estructura de dicho registro se puede ver a continuación:

	Ciudad	Estudios	Actividad	Sexo	T. Inicio	T. Final	Nº Pista	Nº Cinta
1	Castro	Univ.	Profesor	M	00:00	01:13	1	1
2	Pta. Arenas	Medio	Op. Telef.	F	01:13	02:34	1	1
3	Pta. Arenas	Medio	Chofer	M	02:34	03:46	1	1
4	Pta. Arenas	Medio	Chofer	M	03:46	04:30	1	1
5	Pto. Varas	Univ.	Contador	F	04:30	05:38	1	1
6	C. de Velez	Medio	D. Casa	F	05:38	07:19	1	1
7	C. de Velez	Medio	Chofer	M	07:19	09:13	1	1

Tabla 5.1 – Planilla Excel donde fueron registradas las grabaciones de la Base de Datos. (Fuente: Elaboración propia).

5.5 Edición

Una vez finalizado el proceso de grabación de todas las señales que formarán parte de la Base de Datos, fue necesario recurrir a los estudios de grabación de nuestra universidad, para traspasar la información digital contenida en los videocasetes de formato S-VHS al formato de audio convencional para computadores WAV. Este formato contiene como información la representación digital de la forma de onda del sonido. Esto se hace con el fin de tener toda la información en ambiente computacional, ya que es el ambiente en que la Base de Datos trabajará.

Como resultado de éste traspaso, se obtienen 12 archivos WAV de 40 minutos cada uno aproximadamente, que corresponde a la duración de una pista del videocasete S-VHS. Cada archivo WAV contiene como promedio 30 personas grabadas en forma consecutiva, por lo que con ayuda del programa de edición de audio Sound Forge, se procede a identificar con un número, y a separar la grabación de cada locutor de estos archivos wav de 40 minutos, de acuerdo al

registro que se tiene en planilla Excel de los tiempos de inicio y término de todas las grabaciones realizadas.

El nuevo archivo wav resultante fue separado a la vez en las mismas 5 partes en que está dividido el instructivo de las grabaciones, donde a cada archivo que resulte se le identificará con el nombre de Identificación, Vocales, Números, Frases o Prosodia según corresponda la información.

5.5.1 Normalización de los Archivos WAV

Al momento en que se comenzó a editar los archivos WAV, se constató que éstos tenían un nivel sonoro muy bajo, siendo casi imperceptibles tanto visualmente en el editor de audio como auditivamente; por lo que se optó por aplicar el proceso de Normalización sobre todas las muestras de la Base de Datos en forma individual.

Al Normalizar un archivo, se aumenta la intensidad del sonido, de modo que el nivel más alto de la muestra alcance un nivel definido (Umbral) sin distorsionarlo y de una manera proporcional a cada zona de la curva envolvente del sonido que está formada por la sucesión de las amplitudes de cada muestra individual del sonido. Con este proceso se asegura el utilizar todo el rango dinámico disponible sin llegar a la saturación ni afectando a la calidad de la onda.

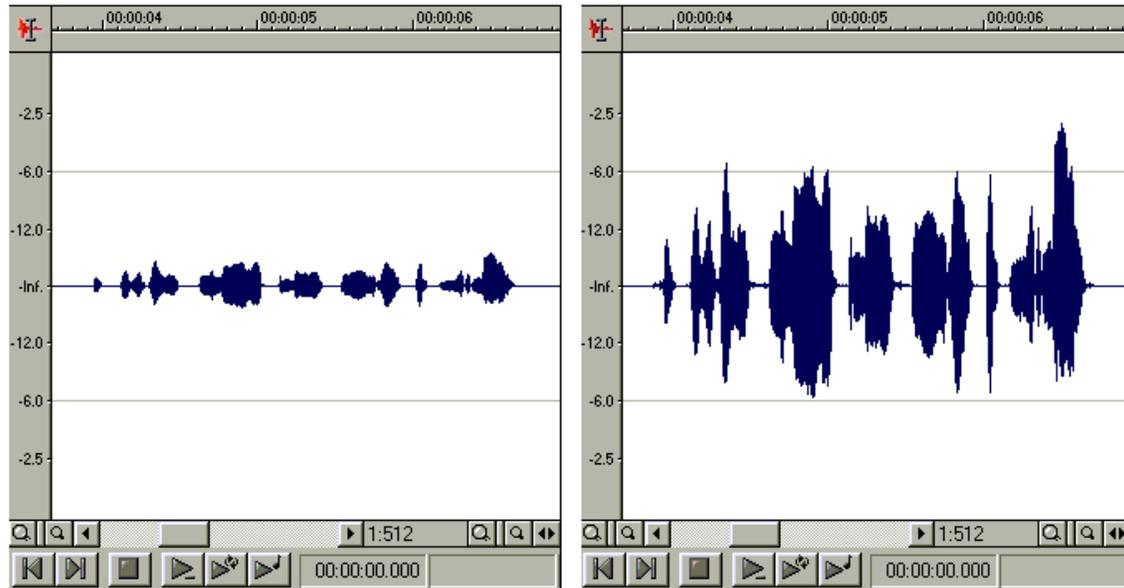


Fig. 5.3 - Señal de audio antes y después del proceso de Normalización. (Fuente de Elaboración propia).

Se estableció el umbral de normalización en -3dB , común para todos los archivos con el fin que todas las señales fueran perceptibles auditivamente y favorecerlas al momento de ser analizadas en los programas especializados que se utilizarán posteriormente, debido a que la señal es asimilada de mejor manera pues su rango dinámico ha sido optimizado, pudiendo visualizar y extraer los parámetros de interés; al contrario de una señal que no estuviese normalizada.

5.5.2 Función de transferencia del Sistema telefónico implementado

Tomando en cuenta los antecedentes que se tienen con respecto a la variación que produce el accionar del Híbrido Telefónico sobre la voz humana, fue necesario evaluar su *Respuesta de Frecuencia* y *Función de Transferencia*.

La Función de Transferencia del Híbrido se pudo medir con la ayuda de un archivo en formato WAV estéreo que se registró en el momento de determinar las variaciones del híbrido sobre la voz humana. Esto se hizo utilizando los dos

sistemas electroacústicos implementados, los cuales se someten a un barrido de frecuencias que va desde los 80 Hz a los 20 KHz, obteniendo el archivo WAV con la respuesta de frecuencia de cada sistema, el canal izquierdo del archivo posee la información del barrido según la respuesta del híbrido y el otro la información según la respuesta del micrófono, esta última es la referencia ya que es un registro exacto de todas las frecuencias a las que fue sometido. Las respuestas de frecuencias obtenidas en esta sección fueron:

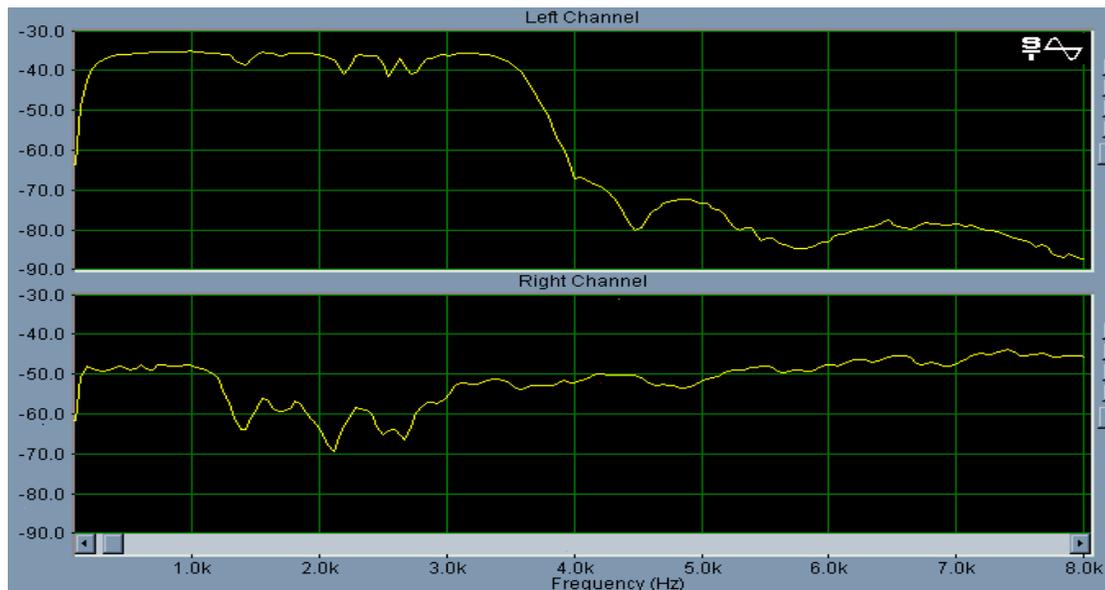


Fig. 5.4 – Respuestas de Frecuencia del híbrido telefónico (Left Channel), Respuesta de Frecuencia del micrófono (Right Channel). (Fuente: Elaboración propia).

La función de Transferencia nos permitirá determinar como actúa el Híbrido Telefónico para este rango de frecuencias, esto se determinará a partir de la señal de salida del híbrido con respecto a la señal que ingresa al dispositivo, que en este caso es la señal registrada por el micrófono.

Teóricamente ésta es la razón de la señal que sale del dispositivo con la señal que entra, dicha evaluación fue realizada con la ayuda del programa SpectraLab, utilizando la función “Real Transfer Function R/L” sobre el archivo WAV estéreo, obteniendo los siguientes resultados:



Fig. 5.5 – *Función de Transferencia del híbrido telefónico. (Fuente: Elaboración propia)*

Como resultado de la Función de Transferencia del Híbrido y su Respuesta de Frecuencia, es posible determinar que el híbrido no responde aproximadamente sobre los 3.5 KHz.

Por lo tanto, se tomó la determinación se remuestrear todos los archivos de la Base de Datos, debido a que éstos se encontraban muestreados a una frecuencia de 48 KHz. Debido al teorema de Niquist, el cual plantea que la frecuencia mínima de muestreo debe ser el doble del ancho de banda de la señal original y como nuestro sistema no responde aproximadamente por sobre los 3.5 KHz, se remuestrearon los archivos a una frecuencia de 8 KHz.

A modo de observación se puede decir que ésta determinación se valida con el hecho de que Bases de Datos tan importantes como la *Telvoice* para reconocimiento de personas realizado en España (detallada en el Capítulo 3) lo tomaron en cuenta y consideraron frecuencias de muestreo KHz.

El Remuestreo de información cambia el número de muestras grabadas por segundo. Al convertirse en una tasa de muestra superior, las muestras adicionales son interpoladas y el tamaño del archivo aumenta. Al convertirse en una tasa de muestra inferior, algunas pruebas son removidas y el tamaño del archivo decrece.

La información que se pierde por sobre los 4 Khz en cada señal, no es relevante para el desarrollo de los distintos estudios que se pueden llevar a cabo. Las diferencias obtenidas por el resamplero de las señales son de 0.1 %, como se puede apreciar en la siguiente tabla:

	Vocales	Fo (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Shim(%)	Jitt(%)	NHR(%)
48 KHz	/a/	119.0	716.0	1225.0	2523.0	5.0	0.48	17.5
	/e/	119.0	428.0	2094.0	2594.0			
	/i/	121.0	301.0	2264.0	2756.0			
	/o/	121.0	476.0	840.0	2648.0			
	/u/	124.0	364.0	744.0	2389.0			
8 KHz	/a/	119.0	716.4	1225.0	2522.0	5.0	0.47	17.5
	/e/	119.0	428.0	2094.0	2594.0			
	/i/	121.0	301.0	2264.0	2756.0			
	/o/	121.0	476.0	840.0	2647.0			
	/u/	124.0	364.0	744.0	2387.0			

Tabla 5.2 – Tabla comparativa de las frecuencias vocálicas muestradas a 48 Khz y a 8 Khz. y los descriptores acústicos y la vocal **a**. (Fuente: Elaboración propia).

Una de las grandes ventajas producto del remuestreo de todos los archivos, es la disminución de la capacidad de almacenamiento en formato WAV de las muestras, la cual disminuyó en un 80 %, por lo que la información de toda la Base de Datos que estaba almacenada en 3 discos compactos se logro traspasar sólo a uno, tomando en cuenta que este remuestreo afecta de forma mínima a las muestras.

5.5.3 Resultados comparativos de mediciones Micrófono-Teléfono.

Como se mencionó en el ítem **5.1.1** de esta sección, la evaluación de la respuesta del híbrido telefónico se hizo comparando los registros de cuatro personas hechos por los dos sistemas electroacústicos implementados. Una vez obtenida la información de los distintos parámetros de interés, se comienza la comparación, por persona, de los valores extraídos de los distintos medios de grabación, de lo cual se obtienen los siguientes resultados:

	Vocales	Fo (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Shim(%)	Jitt(%)	NHR(%)
Micrófono	<i>/a/</i>	119.0	684.0	1228.0	2732.0	2.5	0.45	26.0
	<i>/e/</i>	119.0	337.0	2144.0	2958.0			
	<i>/i/</i>	121.0	244.0	2311.0	3253.0			
	<i>/o/</i>	121.0	352.0	739.0	2851.0			
	<i>/u/</i>	124.0	281.0	678.0	2697.0			
Teléfono	<i>/a/</i>	119.0	716.0	1225.0	2523.0	5.0	0.48	17.5
	<i>/e/</i>	119.0	428.0	2094.0	2594.0			
	<i>/i/</i>	121.0	301.0	2264.0	2756.0			
	<i>/o/</i>	121.0	476.0	840.0	2648.0			
	<i>/u/</i>	124.0	364.0	744.0	2389.0			

Tabla 5.3- Variaciones de frecuencia obtenidos por el registro en distintos medios para una persona. (Fuente: Elaboración propia).

La tabla 5.3 expuesta anteriormente, corresponde sólo a uno de los cuatro locutores grabados con el fin de estimar el comportamiento del híbrido telefónico sobre la señal de voz, las 3 tablas restantes obtenidas se encuentran en los anexos de esta tesis. Se puede apreciar que las variaciones que ocurren debido al comportamiento del híbrido, no son las mismas para todos los parámetros evaluados, sino que existen parámetros que están más expuestos a sufrir

variaciones como el *SHIMMER* con un 55% y *NHR* con un 18%, mientras que otro parámetro susceptible a variar es la primera formante en un 10% con respecto a la señal original.

La variación porcentual de los valores obtenidos entre la señal grabada por micrófono y la grabada por medio del híbrido telefónico a nivel general, es del 6,3%, donde se evaluaron los promedios de variación de cada uno de los cuatro locutores registrados.

También se evalúa la variación que sufre la amplitud de los formantes con respecto a la amplitud de la frecuencia fundamental de cada persona según los archivos registrados por medio del híbrido telefónico o por medio del micrófono. Los resultados obtenidos son los siguientes:

		dF1(dB)	dF2(dB)	dF3(dB)
Micrófono	<i>/a/</i>	-9.3	-15.0	-37.0
	<i>/e/</i>	-5.6	-29.0	-37.0
	<i>/i/</i>	0.7	-37.0	-34.0
	<i>/o/</i>	-8.0	-17.0	-46.0
	<i>/u/</i>	-2.4	-16.0	-60.0
Teléfono	<i>/a/</i>	30.0	25.0	4.3
	<i>/e/</i>	27.0	20.0	12.0
	<i>/i/</i>	26.0	11.7	5.0
	<i>/o/</i>	30.0	24.0	0.9
	<i>/u/</i>	22.0	16.6	-21.0

Tabla 5.4. Variación en decibeles de cada formante con respecto a la amplitud de la frecuencia fundamental para una persona, según los distintos medios de grabación. (Fuente: Elaboración propia).

Como se puede apreciar en la tabla 5.4, las variaciones de las amplitudes de cada formante son muy grandes para los distintos medios de grabación de acuerdo al valor de referencia de la frecuencia fundamental para cada archivo analizado; esto se solventa con el porcentaje de variación global para cada formante que se muestra en la siguiente tabla:

	dF1	dF2	dF3
/a/	91%	68%	37%
/e/	93%	51%	40%
/i/	104%	53%	26%
/o/	97%	69%	16%
/u/	99%	65%	28%

Tabla 5.5 Variación porcentual de la amplitud de cada formante con respecto a la frecuencia fundamental para los distintos medios de grabación. (Fuente: Elaboración propia).

De la tabla 5.5, se desprende que la amplitud de los formantes tiene un porcentaje total de variación del 62%, para los dos sistemas de grabación; Lo cual si bien es alto, no afecta nuestro análisis ya que el valor de las formantes, armónicos y de la frecuencia fundamental no varía en el eje de las frecuencias; pero si hay un aumento en la Amplitud no solo de las formantes, sino de toda la información de cada señal registrada vía híbrido telefónico; esto se puede traducir como la amplificación de solo un rango de frecuencias que es intrínseco del sistema utilizado para realizar los registros de la Base de Datos.

5.6 Buscador Inteligente de la Base de Datos

Se denominó “Buscador Inteligente” de la Base de Datos al programa que permite una herramienta de fácil uso y rápido acceso a cualquier registro sonoro que pertenezca a ésta. Es un programa creado bajo ambiente Visual Basic, por lo que su aplicación puede ser instalada en el disco duro de cualquier computador para luego leer e importar los registros sonoros con el fin de realizar análisis de estos según se estime conveniente por el usuario, sin modificar los archivos originales que se encuentran ordenadamente almacenados en un disco compacto.

Posee regiones de búsqueda por Zona del país, Ciudad, Sexo, Estudios y actividad de la o las personas que se desean buscar para evaluar. Esta región de búsqueda se encuentra en el costado izquierdo de la ventana principal del buscador, donde una vez que se escogieron los parámetros de búsqueda, se procede a presionar el botón “Buscar” para que el buscador comience a trabajar sobre las muestras de toda la Base; los resultados de la búsqueda aparecerán en la misma ventana principal, en la cual por medio de un click con el mouse sobre la persona que se desea examinar, aparece una ventana secundaria con las opciones de escuchar dicho registro o copiarlo al disco para posterior evaluación y análisis. Los archivos que se pueden copiar u oír son los registrados por cada persona como las vocales, frases, números y prosodia.

5.7 Metodologías

5.7.1 Programas de Análisis

A continuación se hará una reseña de los programas computacionales especializados utilizados para el desarrollo de los distintos tipos de estudio a los que la Base de Datos está enfocada. Estos fueron:

- ✓ SpectraLab
- ✓ Praat

5.7.1.1 SpectraLab

Es un poderoso analizador de espectros que puede ser utilizado para señales mono y estéreo; el preciso análisis espectral que realiza, que es el motivo por el cual fue utilizado, puede ser hecho en tiempo real, o con señales previamente grabadas como las que posee la Base de Datos.

El programa SpectraLab tiene opciones de visualización sobre el archivo cargado que pueden ser escogidas en el menú “View”, estas opciones son:

- ✓ **Time Series:** Muestra la forma de onda del audio digitalizado. Es similar a un despliegue del osciloscopio, es decir, nos muestra la amplitud versus el tiempo.
- ✓ **Spectrum:** Muestra la amplitud de la señal versus la frecuencia.
- ✓ **Phase:** Muestra la fase de la señal versus la frecuencia.
- ✓ **Spectrogram:** Muestra el espectrograma, o sea, los espectros con el paso del tiempo. La amplitud de la señal es mostrada según color o escala de grises.
- ✓ **3-D Surface:** Muestra los espectros con el paso del tiempo en 3-D.

Para analizar el espectro de la señal deseada, se sigue el procedimiento que se explica a continuación, una vez cargado el archivo:

En el Menú “View” se escoge la opción “*Time Series*”, luego se procede a escoger la herramienta de selección gráfica  que se encuentra la esquina superior izquierda de esta ventana, luego se procede a seleccionar manualmente el trozo más estable en amplitud de la señal que se analizará, haciendo clic y arrastrando el mouse a través del rango de información deseada.

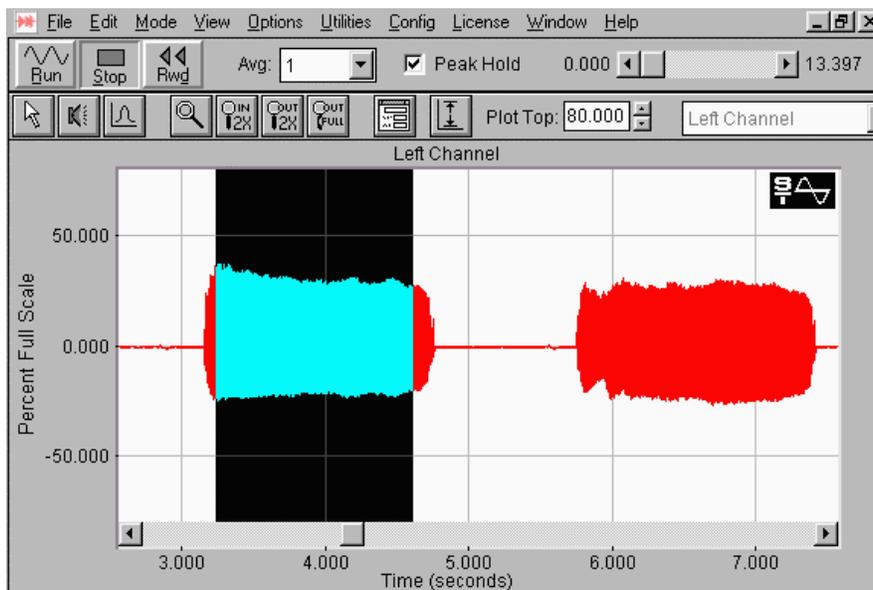


Fig. 5.6 - Time Series programa SpectraLab. (Fuente: Elaboración propia).

Una vez hecha la selección del trozo deseado, se realiza un clic en el botón derecho del mouse, desplegándose el siguiente menú de opciones:

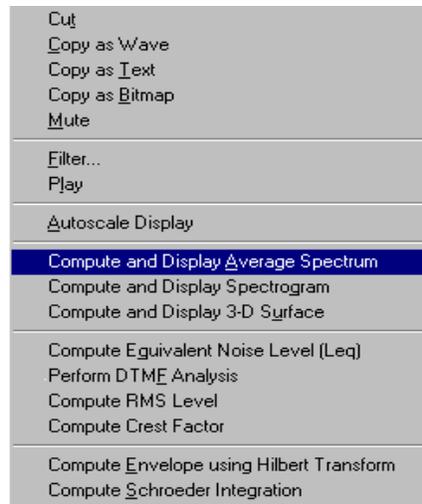


Fig. 5.7 – Display opciones time series programa SpectraLab. (Fuente: Elaboración propia).

Este Menú, como se puede ver, posee varias opciones de cálculo, pero para nuestro trabajo nos limitamos a escoger la opción “*Compute and Display Average Spectrum*”, el cual nos calculará y mostrará un espectro promedio de la región seleccionada, como se muestra a continuación:

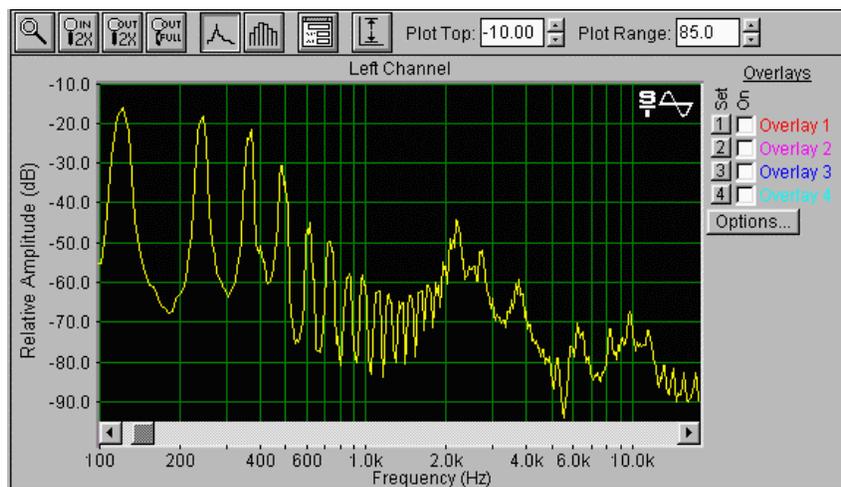


Fig. 5.8 - Display Average Spectrum programa SpectraLab. (Fuente: Elaboración propia).

Se debe dejar en claro que la forma de cálculo del espectro, promediado y discreto, es sólo una de las muchas formas existentes para hacerlo con éste programa, pero es la que sirvió para el progreso de éste trabajo.

5.7.1.2 Praat

Es un programa computacional principalmente utilizado por especialistas fonéticos, para editar y analizar señales del habla humana. La utilización de este programa fonético acústico es de interés debido a que de él se obtienen los cálculos, análisis y gráficos de los parámetros que competen a este trabajo de tesis. Los cálculos que Praat es capaz de realizar son variados dentro del área de la fonética, pero serán acotados según las necesidades demandadas:

- ✓ Análisis de Formantes
- ✓ Análisis de Voz
- ✓ Análisis del tono fundamental (Pitch)

Todos los análisis en Praat se hacen a partir del archivo de audio cargado que se desea analizar, la forma de cargar un archivo en Praat, es como se muestra a continuación:

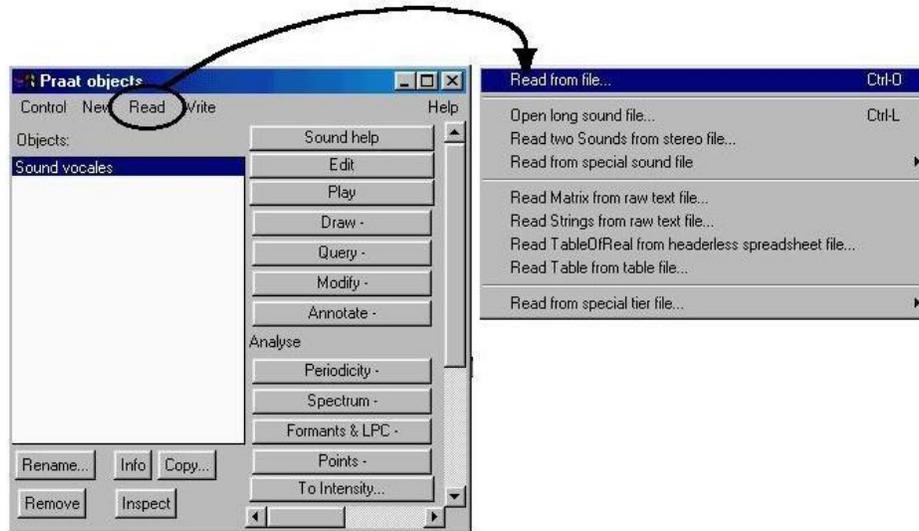


Fig. 5.9 – Menú Read programa Praat. (Fuente: Elaboración propia).

Seleccionando el Menu “Read” se despliega un sub-menú. Donde se debe elegir el comando “Read from file...”, aquí se puede acceder a la ruta del archivo que se desea analizar. Una vez que el nombre del archivo aparece en la ventana “Objects” es porque éste ha sido cargado.

Para acceder a los distintos tipos de análisis que puede hacer éste programa, se presiona el botón “Edit”, que se encuentra en la misma ventana principal del programa, lo que mostrará la siguiente ventana:

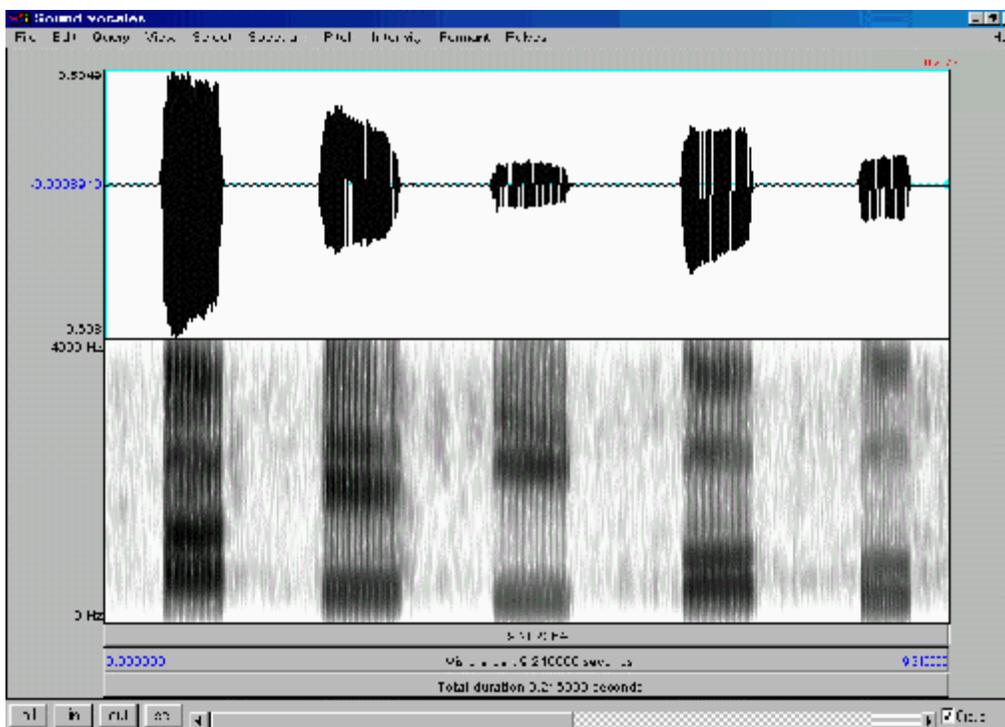


Fig. 5.10 – Ventana de edición al presionar el botón *Edit* del menú principal del programa Praat. (Fuente: Elaboración propia.)

Esta ventana de edición muestra la forma de onda de todo el archivo en conjunto con su espectrograma, donde se pueden ver las respectivas amplitudes, frecuencias y tiempos. El espectrograma permite ver el contenido espectral de la señal en función del tiempo, donde el eje horizontal corresponde al tiempo, mientras que el eje vertical representa la frecuencia; también se puede apreciar en esta ventana los menús que serán utilizados, como “*Spectrum*”, “*Pitch*”, “*Formant*” y “*Pulses*”.

5.7.1.2.1 Análisis de Formantes

Este análisis se hace a partir de la ventana de edición del archivo, donde seleccionando el menú “*Formant*” y el comando “*Show formants*”, se muestran unas líneas rojas sobre el área del espectrograma, que coinciden con la zona más

oscura del espectrograma, estas líneas rojas corresponden a los contornos de las formantes de la señal. Se pueden obtener valores discretos o promediados de las formantes; un valor discreto se obtendría solo pinchando con el mouse sobre el área más oscurecida del espectrograma, mientras que valores promediados se pueden obtener seleccionando un trozo del archivo y presionando las teclas F1, F2, F3, según la formante que se desee conocer:

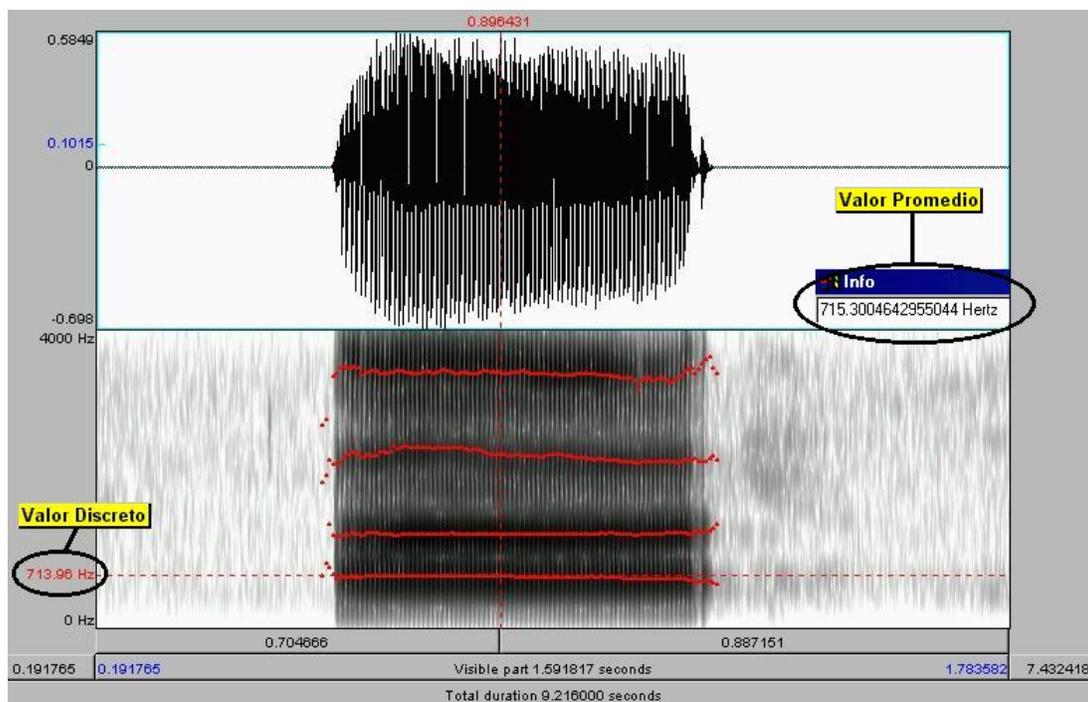


Fig. 5.11 – Ventana de edición con selección del menú *Show formants* del programa Praat.
(Fuente: Elaboración propia).

El comando “Formant settings” de este mismo menú, posee opciones las cuales variaran los valores de cálculo de las formantes y su visualización según se desee.

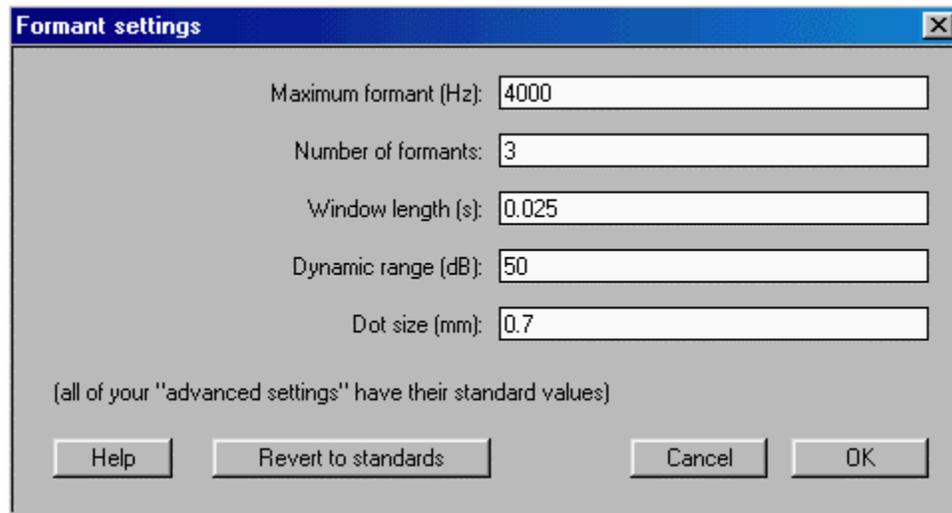


Fig. 5.12 – Comando Formant settings del programa Praat. (Fuente: Elaboración propia).

- ✓ **Maximun formant (Hz.):** Es el rango máximo de búsqueda de las formantes.
- ✓ **Number of Formants:** Número de formantes que se encontrarán.
- ✓ **Window length (s):** Duración de la ventana que se utilizará en el análisis.
- ✓ **Dynamic range (dB):** El rango dinámico de la señal que tomará el programa para analizar.

OBS: Para nuestros análisis es necesario ocupar un rango máximo de formantes de 4000 Hz., así como también la cantidad de 3 formantes debido a la frecuencia de muestreo del archivo de audio.

5.7.1.2.2 Análisis de Voz

En la misma ventana de edición del archivo, es posible el análisis de voz a partir de los pulsos glotales, los cuales son visibles como líneas azules verticales a

través de la forma de onda de la señal, la forma de activar estos pulsos es seleccionando en el menú “*Pulses*”, la opción “*Show Pulses*”:

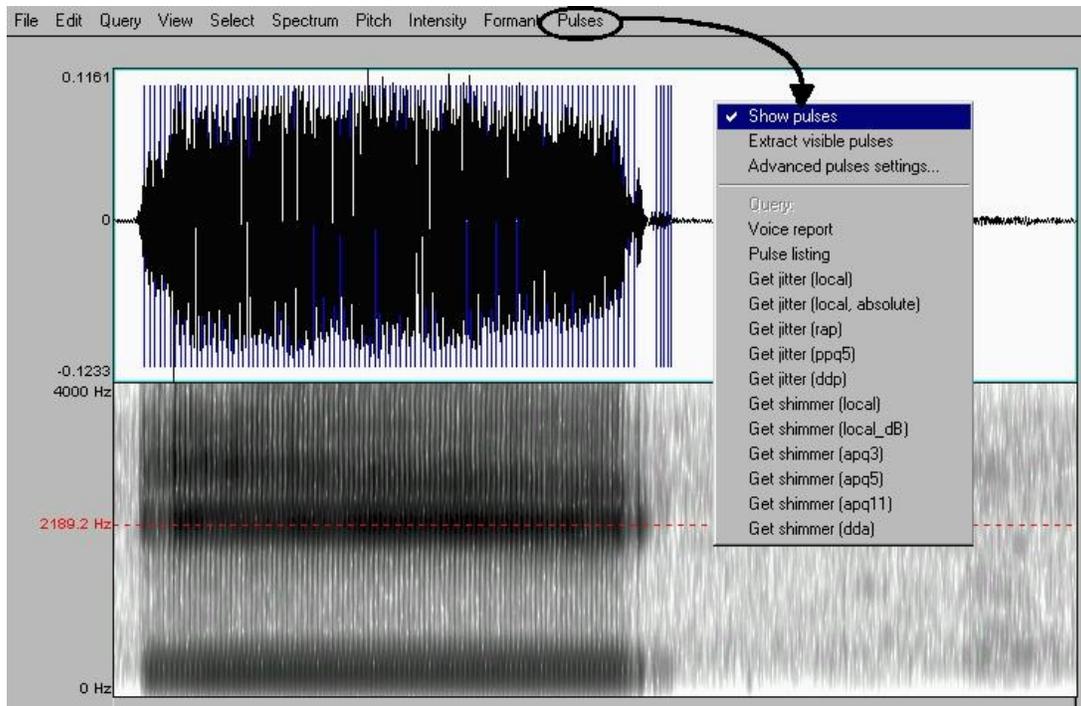


Fig. 5.13– Ventana de edición con selección del menú *Pulses* del programa Praat. (Fuente: *Elaboración propia*).

El menú “*Pulses*” contiene el comando “*Voice Report*”, que una ventana distinta entregará los resultados medidos de parámetros de la voz para una parte del sonido seleccionado previamente como el SHIMMER, JITTER y quiebres de VOZ.

5.7.1.2.3 Análisis del tono fundamental (“*Pitch*”)

Para realizar este análisis se debe visualizar la variación de la frecuencia fundamental a lo largo del tiempo en una señal cargada, lo cual se selecciona el menú “*Pitch*” y luego el comando “*Show Pitch*”:

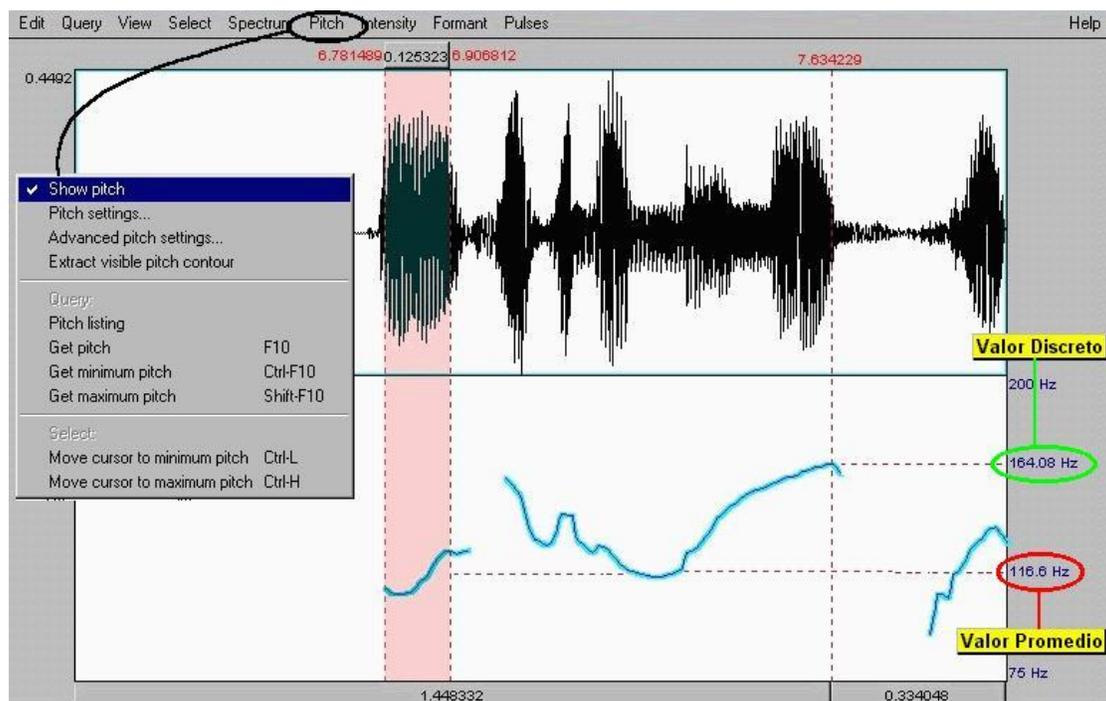


Fig. 5.13– Ventana de edición con selección del menú *Pitch* del programa Praat. (Fuente: *Elaboración propia*).

Se debe notar que la mitad de abajo de esta ventana contiene el contorno de pitch, dibujado como una línea azul, ésta es la variación del tono fundamental a lo largo del tiempo, a la derecha de la ventana se ven 4 valores escritos con dígitos azules, los dos valores que se encuentran en los extremos superior e inferior determinan el rango de la frecuencia fundamental que mostrará el programa, los otros dos valores, son valores de la frecuencia fundamental evaluados para distintos casos; El valor discreto es el valor de la fundamental en

la posición del cursor, mientras que el valor promedio se obtiene a partir de una región seleccionada.

5.7.2 División de muestras a Evaluar

Para el desarrollo de los distintos tipos de estudios que se pueden realizar con la Base de Datos, se utilizó el 10 % del total de muestras con que cuenta la Base de Datos, el cual corresponde a 31 muestras tomadas al azar. La estructura de los estudios quedó distribuido de la siguiente manera:

- ✓ Estudio del tono medio hablado: 8 personas
- ✓ Estudio de Voz Normal y Patológica: 12 Personas
- ✓ Estudio de la Prosodia: 3 Personas
- ✓ Estudio de Fonemas: 8 Personas

5.7.2.1 Evaluación del Tono Medio Hablado

La evaluación del Tono Medio Hablado de la persona, se realiza imitando los pasos explicados anteriormente para el análisis del tono fundamental.

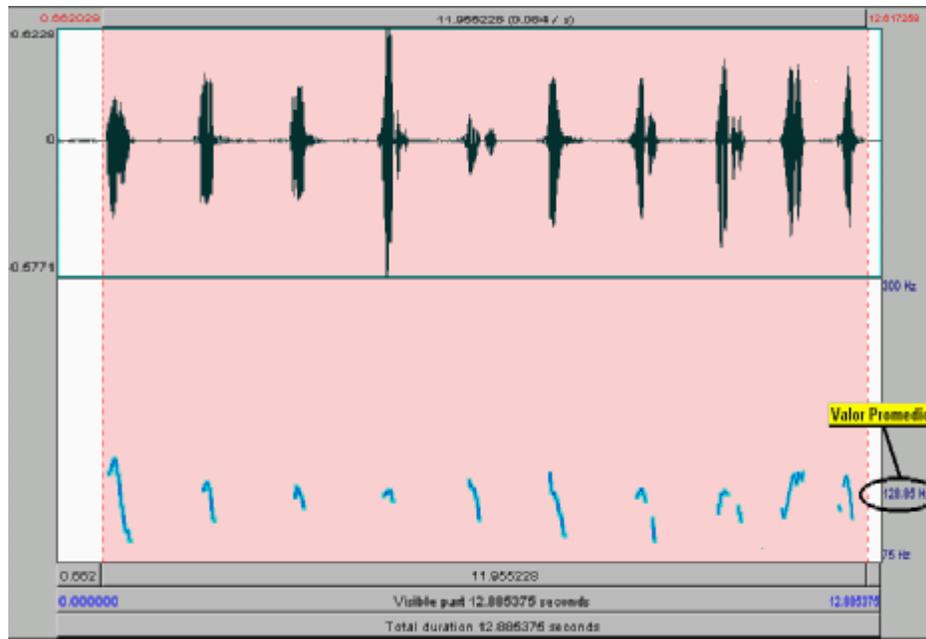


Fig. 5.14 – Serie automática de números (Parte superior) y la representación de los contornos de Pitch (parte inferior) en el programa Praat. (Fuente: Elaboración propia).

Cargando, en el programa Praat, el archivo que contiene la serie de números que ejecutó cada persona; se selecciona el rango del archivo que contenga la información de audio de la serie automática; lo que entrega, en el costado derecho de la pantalla de Pitch, el valor promedio de la frecuencia fundamental, el cual se interpreta como el tono medio hablado de la persona.

5.7.2.2 Evaluación de la Carta de Formantes

Para la evaluación y obtención de la carta de formantes se trabaja con el archivo que contenga las vocales, el procedimiento que se muestra a continuación, se realizó por cada vocal con el objeto de obtener menos errores en el cálculo de la carta de formante de cada persona. El proceso comienza editando el archivo completo que contiene la información vocálica, de la siguiente forma:

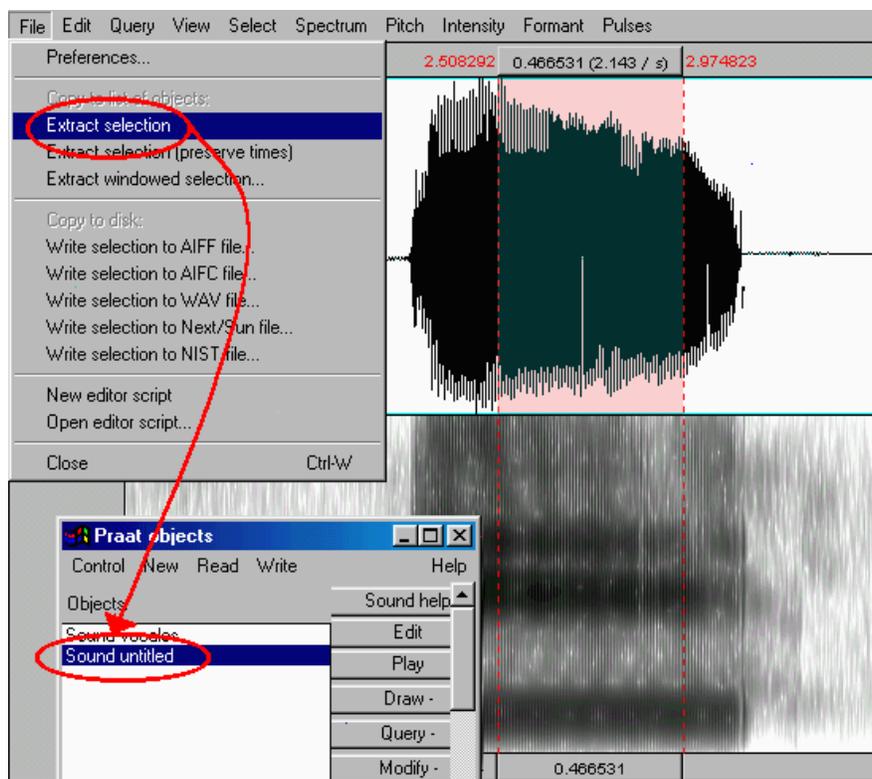


Fig. 5.15– Ventana de edición con selección del menú *File* y el comando *Extract Selection* del programa Praat. (Fuente: Elaboración propia).

Se selecciona la parte más estable de la vocal en cuanto a su amplitud, luego se extrae esa selección de audio, mediante el Menú “*File*” y el comando “*Extract Selection*”, esto nos crea un archivo nuevo llamado “*Sound Untitled*” que contiene la información de la parte seleccionada, el cual se visualiza en la ventana principal del programa en la sección “*Objects*”.

Mediante el comando “*Rename*”, se puede cambiar el nombre a este nuevo archivo, sugiriendo que se le denomine con el nombre de cada vocal extraída, para realizar una forma de trabajo más fluida.

A este nuevo archivo se le calculan las formantes por medio del comando “*To Formant (Burg)*” perteneciente al menú “*Formants & LPC*”, donde se despliega un menú de opciones para el cálculo, el cual crea un nuevo archivo con el cálculo

de las formantes, según un análisis espectral a corto tiempo según el número de formantes que se deseen calcular.

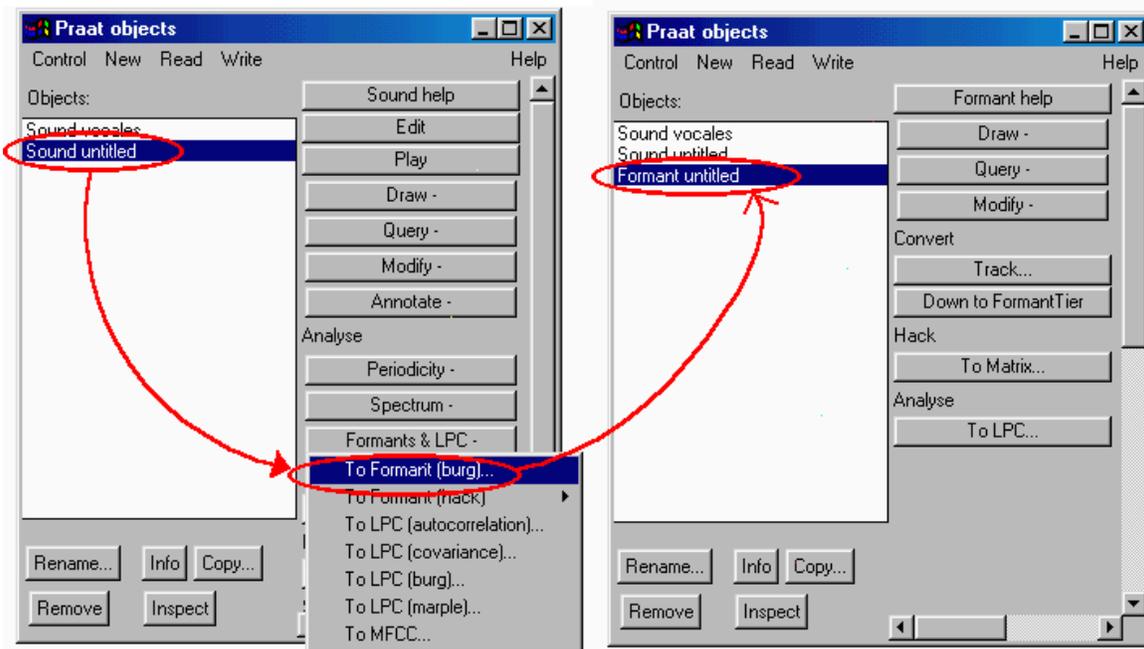


Fig. 5.16 – Comandos para la obtención de las formantes en el programa Praat. (Fuente: Elaboración propia).

Este nuevo archivo aparece de la misma forma que el creado anteriormente, pero su nombre ahora es “*Format Untitled*”; la forma de llevar este archivo a la carta de formantes, es por medio del menú “*Draw*” y su comando “*Scatter plot (reversed axes)*”, de donde se ejecuta el siguiente sub-menú para la configuración visual de las opciones de la carta de formantes que se desea:

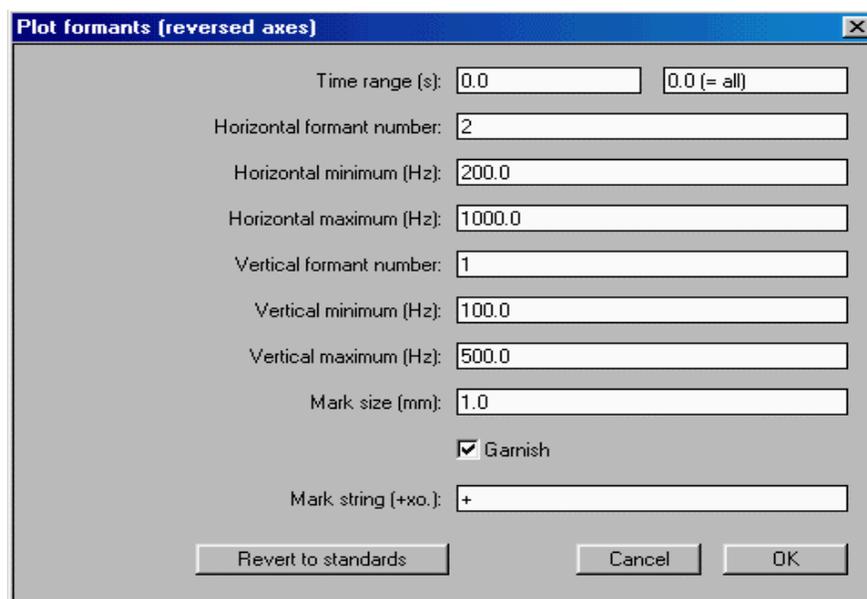


Fig. 5.17 – Comando *Plot Formants* del programa Praat. (Fuente: Elaboración propia).

Al ejecutar este menú, el programa nos muestra la carta de formantes y la ubicación que tiene aquí la vocal analizada. Para realizar la gráfica de otra vocal, se debe seguir el mismo procedimiento, donde se obtienen los siguientes resultados:



Fig. 5.18 – Carta de Formantes de persona analizada en el programa Praat. (Fuente: Elaboración propia).

5.7.2.3 Evaluación de Voz Normal y Voz Patológica

Existen dos formas complementarias de realizar este análisis:

- ✓ Extracción de Parámetros.
- ✓ Comparación de Espectros.

5.7.2.3.1 Extracción de Parámetros

Este consta de la evaluación de parámetros que son capaces de determinar si una voz es normal o patológica, según los umbrales de normalidad que posean dichos parámetros, como la frecuencia fundamental tanto discreta como promedio, Jitter, Shimmer y NHR. Todos estos parámetros se evalúan solo para la vocal "a", ya que es la vocal que se toma como referencia. A esta altura, se sabe el procedimiento a seguir para conseguir los valores de estos parámetros, sin incluir el parámetro NHR. El NHR se obtiene a partir del nuevo archivo creado de la extracción de audio de la vocal "a" en la ventana de edición, como sabemos, este nuevo archivo se encuentra ahora en la ventana principal del programa.

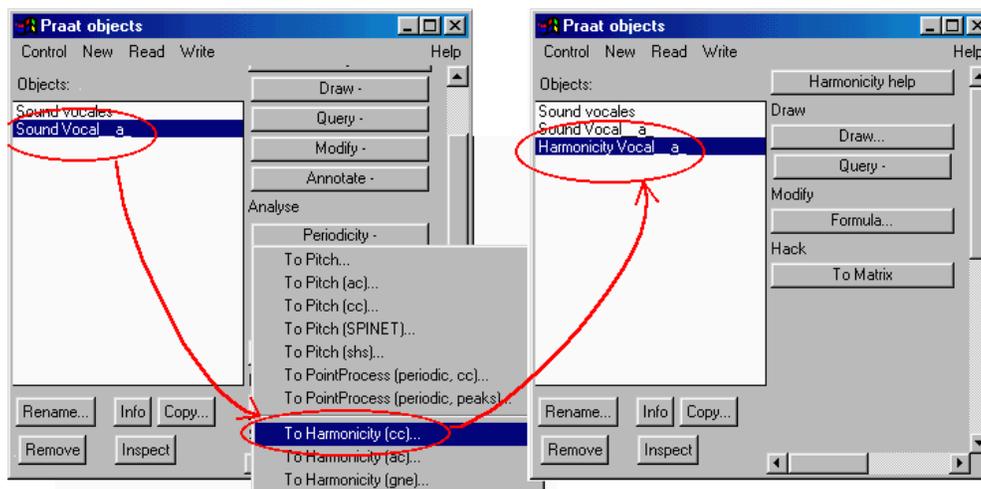


Fig. 5.19 – Comandos para la obtención del NHR en el programa Praat. (Fuente: Elaboración propia).

A este archivo, que denominaremos “Sound Vocal a”, se procede a calcular su NHR, mediante el comando “To Harmonicity(cc)”, perteneciente al menú “Periodicity” del programa; luego de la ventana de configuración para el cálculo de este parámetro, se crea un nuevo archivo con la información del valor del NHR, al cual podemos acceder, mediante el menú “Query” y el comando “Get mean”, entregándonos un valor en decibeles de NHR.

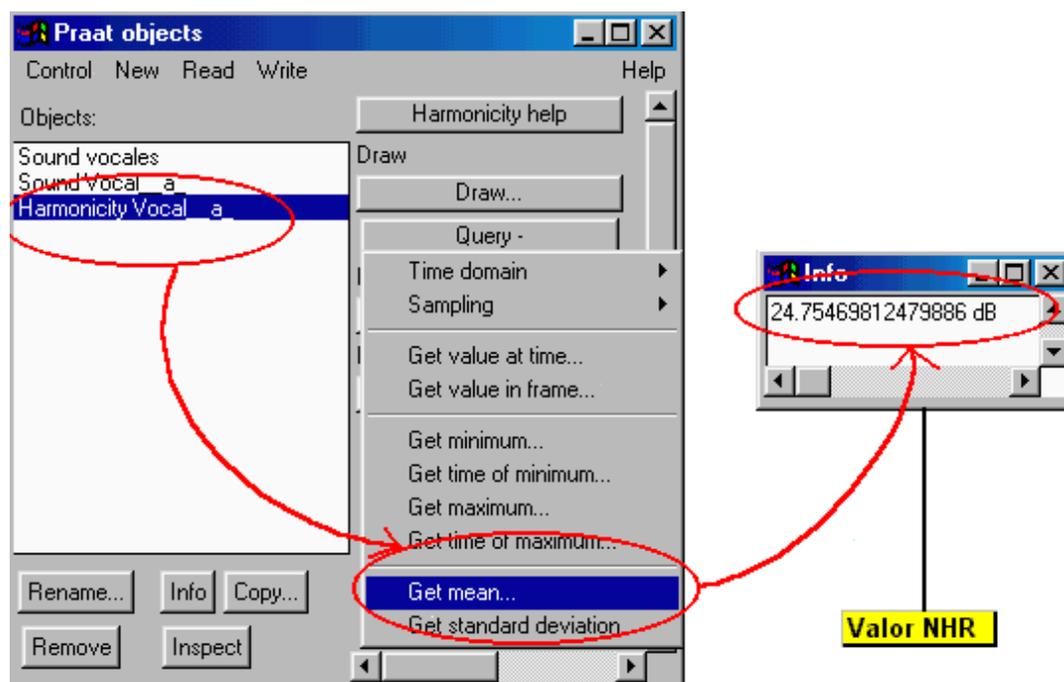


Fig. 5.20 - Menú Query y el comando Get mean del programa Praat. (Elaboración propia).

5.7.2.3.2 Comparación de espectros

La comparación se realiza con la ayuda del espectro de la vocal “a”, para visualizar las características espectrales de la señal, lo que determinará en cierta forma si la persona es normal o patológica de acuerdo a la visualización

clara o no de armónicos, formantes y ruido que pueda poseer la señal. No está demás señalar que la poca claridad del espectro nos permite aseverar que ésta patología del habla puede ser a nivel de cuerdas o a nivel resonancial. Pero también podemos decir que la presencia de ruido en el espectro puede en algunos casos deberse a las características propias del sistema de grabación.

En la figura 5.18 se muestran los espectros de la vocal “a” ocupando el programa SpectraLab, donde el color rojo representa una voz normal y el color blanco una voz patológica.

Este análisis, como se dijo anteriormente, se complementa con los valores de los parámetros obtenidos en Praat.

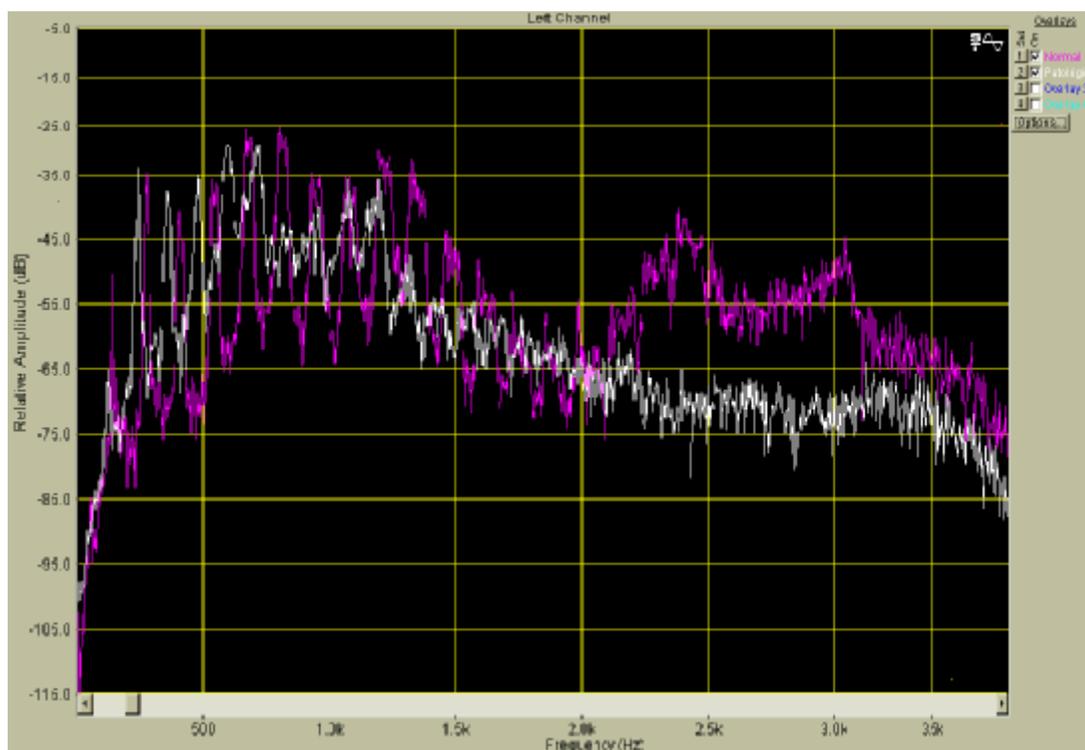


Fig. 5.21 – Espectros de una voz normal y una voz patológica. (Fuente: Elaboración propia).

5.7.2.4 Evaluación de Fonemas

Para el proceso de evaluación de los fonemas se escogieron 3 consonantes dentro del archivo que contiene registrada las frases, a estas se les analiza la estructura formántica que poseen para luego compararlas con otros estudios realizados. No se analizaron los fonemas vocálicos debido a que dicha evaluación forma parte del cálculo de la carta de formantes. Las 3 consonantes elegidas son las siguientes:

<i>/l/</i>	<i>Lateral Sonora</i>
<i>/n/</i>	<i>Nasal Sonora</i>
<i>/s/</i>	<i>Fricativa Sorda</i>

Tabla 5.6 – Fonemas Consonánticos elegidas para el análisis. (Fuente: Elaboración propia).

La evaluación del fonema /l/, se hizo de acuerdo a su posición silábica dentro de la palabra que se analiza, es decir, si hay una vocal antes y/o después de este fonema, encontrando 3 posiciones:

<i>Posición Vocal(-)</i>	<i>Palabra Analizada</i>
<i>/l/-</i>	<i>La</i>
<i>-l/-</i>	<i>Metropolitana</i>
<i>-l/</i>	<i>Alguna</i>

Tabla 5.7 – Fonema Consonántico /l/ en tres posiciones. (Fuente: Elaboración propia).

Para las otras 2 consonantes, sin tomar en cuenta la posición silábica, las palabras que se analizaron fueron las siguientes:

	Palabra Analizada
/n/	manera
/s/	padecen

Tabla 5.7 – Fonema Consonántico /n/ y /s/. (Fuente: Elaboración propia).

El procedimiento que se sigue es el mismo que para análisis de formantes, ya que ésta es la información que nos interesa; una vez que logra ubicar el fonema en la ventana de edición, auditivamente, se selecciona la región, para luego encontrar sus formantes presionando las tecla F1, F2, F3 según corresponda el fonema deseado, el valor que se obtiene, es un valor promedio de la selección hecha.

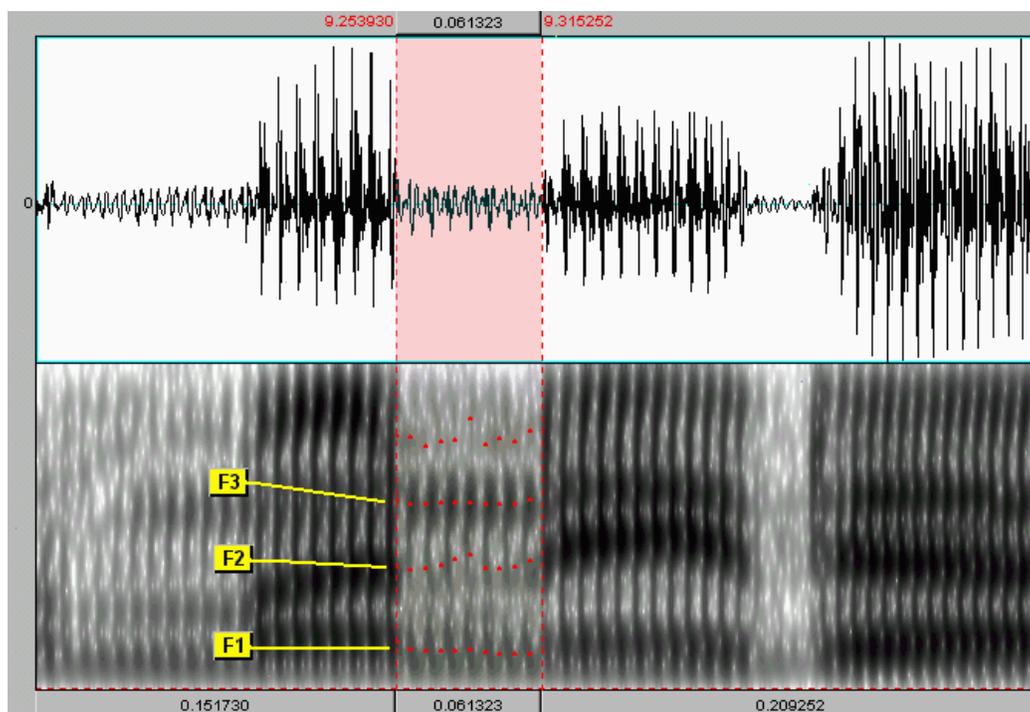


Fig.5.22 - Fonema /n/ con las tres primeras formantes realizadas en el programa Praat. (Fuente: Elaboración propia).

5.7.2.5 Evaluación de Prosodia

La evaluación prosodica se realiza basándose en el análisis de la curva melódica (representación de la evolución temporal de la frecuencia fundamental) obtenida a lo largo del archivo de habla espontánea registrado a cada persona de la Base de Datos.

En este análisis, nos limitaremos a estudiar las diferencias melódicas del acento y la entonación por medio de la estilización de curvas melódicas de una frase relatada espontáneamente por personas de distintas zonas de nuestro país (Norte-Centro-Sur); para esto se estudia a una persona por zona y la frase que se evalúa es la siguiente: *“El Dieciocho de Septiembre se celebra”*.

Para la estilización de las curvas melódicas, se utiliza un algoritmo estilizador que trabaja en ambiente Praat llamado MOMEL (Modeling Melody) el cual es muy utilizado en estudios de representación automática de curvas melódicas. Este algoritmo determina los puntos de inflexión de la curva de la frecuencia fundamental, eliminando variaciones irrelevantes y su funcionamiento no entrega una visualización y cálculo de los segmentos sordos (sin información de F_0) que pertenezcan a la muestra que se evalúa. Una vez cargado el programa estilizador MOMEL, entramos a la siguiente ventana de configuración:

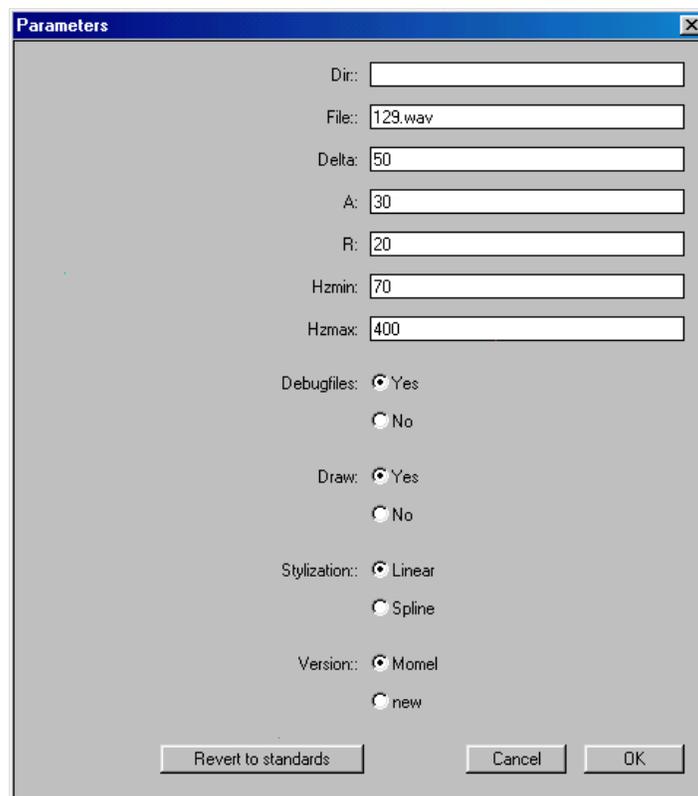


Fig 5.23 - Ventana Principal Algoritmo MOMEL. (Fuente: Elaboración Propia)

La figura 5.23 nos muestra la ventana principal de este estilizador automático, donde se pueden escoger las características de la estilización como el rango de frecuencia fundamental a mostrar, los intervalos con que se tomaran las muestras de F0, porcentaje de error de la estilización de un archivo y el tipo de estilización que se desea. A continuación se mostrarán dos gráficos (Fig. 5.24 y 5.25), donde se puede visualizar claramente una curva estilizada con una no estilizada.

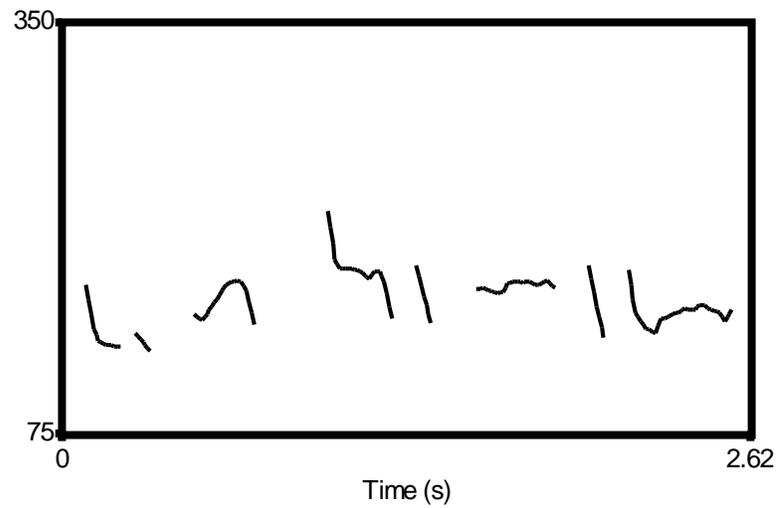


Fig.5.24 - Gráfico obtenido con programa Praat. (Fuente: Elaboración propia).

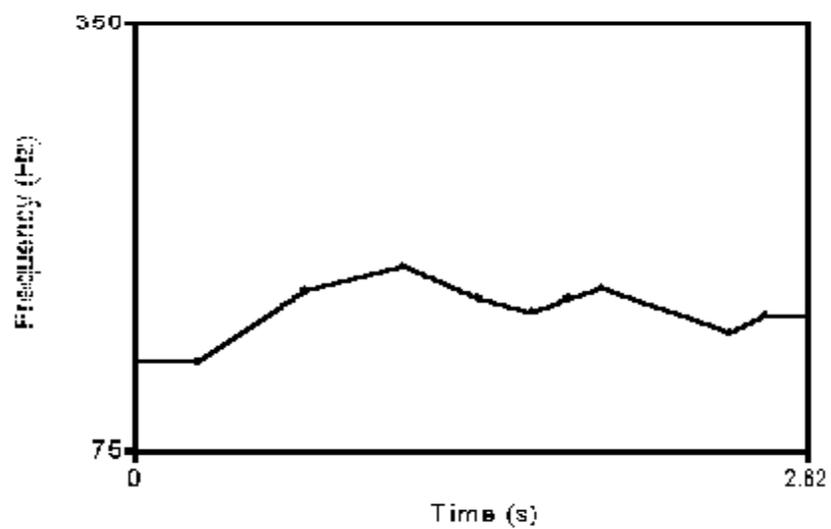


Fig.5.25 - Gráfico obtenido con estilizador MOMEL. (Fuente: Elaboración propia).

6 ANÁLISIS DE RESULTADOS

6.1 Tono medio hablado

En el análisis de resultados para el tono medio hablado de una persona se obtuvieron valores promedio para la Frecuencia Fundamental. A continuación se muestran las tablas conjuntamente con los gráficos del total de muestras, que para éste caso fueron ocho personas, cuatro hombres y cuatro mujeres respectivamente.

Persona	F ₀ Promedio (Hz)
56	141.6
104	162.8
299	142.3
304	130.2

Tabla 6.1 – Tono medio hablado para hablantes masculinos. (Fuente: Elaboración propia).

Persona	F ₀ Promedio (Hz)
10	261.2
57	202.1
34	207.7
62	213.7

Tabla 6.2 – Tono medio hablado para hablantes femeninos. (Fuente: Elaboración propia).

El gráfico asociado a la tabla 6.2, considerando tanto hombres como mujeres, es el siguiente:

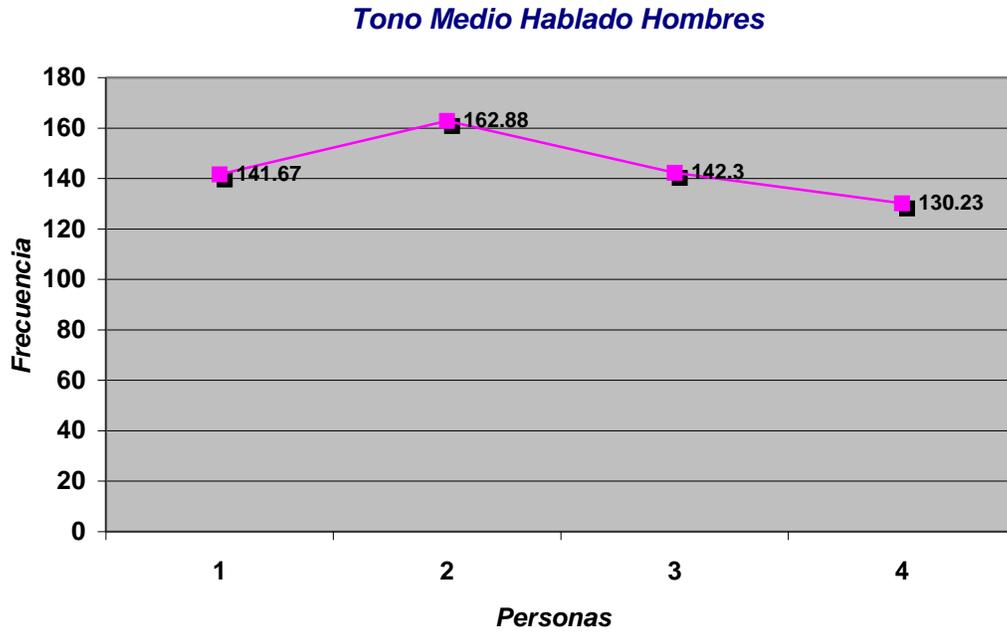


Gráfico 6.1 -Tono medio hablado para hablantes masculinos. (Fuente: Elaboración propia).

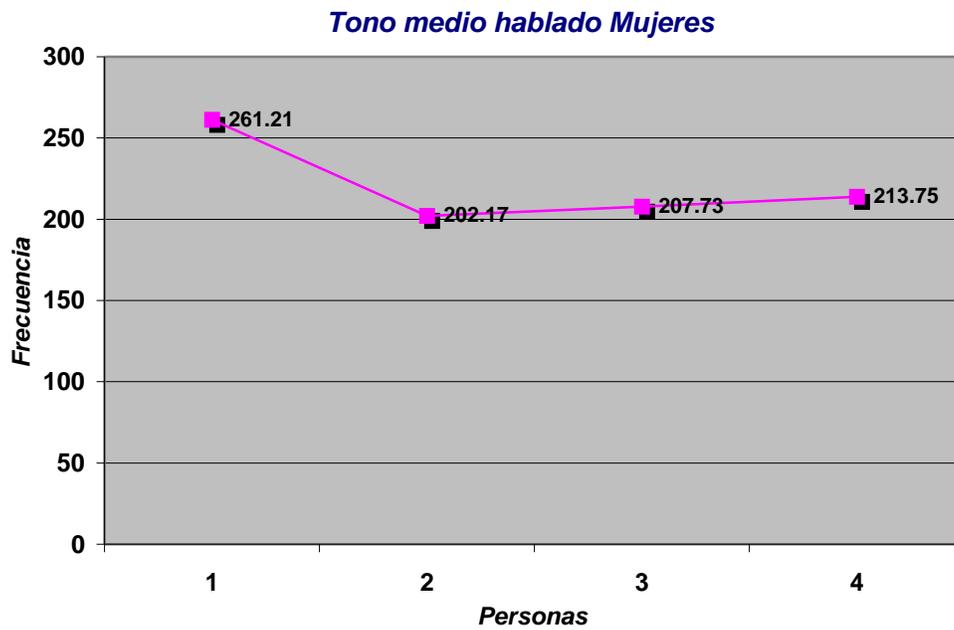


Gráfico 6.2 -Tono medio hablado para hablantes femeninos. (Fuente: Elaboración propia).

En los gráficos 6.1 y 6.2 se aprecian claramente las diferencias de la Frecuencia Fundamental tanto de hombres como de las mujeres. Para el caso de los hombres se observa que no sobrepasan los 200 Hz. mientras que para las mujeres el valor frecuencial es sobre los 200 Hz.

Claro está, que por estructura fisiológica podemos admitir dicha variación y considerar también que es irrelevante la ubicación geográfica del hablante, ya que fueron tomados al azar y sin ningún tipo de restricción.

Si consideramos el estudio realizado por Quilis que se muestra en la tabla 4.1, se puede apreciar que los valores del tono Fundamental obtenidos, tanto para hombres como mujeres, tienen relación y están contenidos en el rango de dicho estudio para cada género.

6.2 Voz Normal y Patológica

6.2.1 Comparación voz normal - patológica

6.2.2.1 Descriptores Acústicos

En base a la selección hecha en el capítulo anterior que tiene relación con las personas, que a juicio de especialistas son las adecuadas para analizar ciertos parámetros acústicos de voz normal y patologías del habla, podemos decir que realizamos una subdivisión de las muestras.

En la siguiente tabla, se muestran a sólo dos hablantes masculinos con las frecuencias formánticas de las vocales y sus respectivos descriptores acústicos.

Persona	Vocales	F0 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Shim(%)	Jitt(%)	NHR(%)
129	/a/	158.1	685.0	1243.0	2501.0	5.4	0.60	12.4
	/e/	158.0	458.0	2017.0	2789.0	11.1	2.20	4.8
	/i/	162.0	403.0	2102.0	2600.4	11.5	0.39	10.1
	/o/	157.0	517.9	982.0	2691.0	13.3	0.89	19.2
	/u/	149.0	439.0	738.1	2493.0	8.8	2.21	18.0
11	/a/	155.3	751.2	1366.4	2513.1	2.8	0.30	25.6
	/e/	149.7	499.0	2998.1	2613.0	3.1	0.20	21.5
	/i/	148.0	352.1	2271.0	2568.3	3.3	0.20	20.0
	/o/	147.1	533.0	890.1	2703.0	1.7	0.30	28.5
	/u/	167.1	398.1	821.3	2603.1	1.8	0.45	24.9

Tabla 6.3 – Tabla comparativa de las frecuencias formánticas de las vocales de una persona con voz normal y otra con voz patológica, además de sus respectivos descriptores acústicos. (Fuente: *Elaboración propia*).

Si se observan los descriptores acústicos de la persona 129 claramente se puede apreciar que el umbral de normalidad se ve alterado.

Recordando los umbrales para estos tres descriptores acústicos, resulta que para la vocal **a** que es la más utilizada en este tipo de análisis, los valores son:

	Teórico	Práctico
Shimmer(%)	3.84	5.81
Jitter(%)	1.04	0.2
NHR(%)	20	12.4

Tabla 6.4 – Tabla comparativa valores prácticos y teóricos.

Sin embargo a modo de observación se puede decir que el hecho de que los análisis se hayan realizado ocupando una señal telefónica, es más probable el margen de error con respecto a los descriptores acústicos, que si se hubiese ocupado otro tipo de señal, como por ejemplo una señal registrada por micrófono.

Por ejemplo, para las perturbaciones de Amplitud (Shimmer) es posible que se produzcan variaciones que no son intencionales al momento de hablar o realizar la grabación, pero sin lugar a duda que ésta variación no significa una tendencia a ser voz patológica, por el contrario. Por lo tanto las variaciones existentes en algunos casos se deben a ésta razón.

6.2.2.2 Espectros

Según los resultados obtenidos fue posible obtener diversos espectros, donde se puede observar claramente el comportamiento espectral de una señal con irregularidades en comparación con una normal.

Se puede inferir que la patología está a nivel de las cuerdas vocales, por lo siguiente:

- ✓ Armónicos poco definidos, se advierte presencia de peaks entre los armónicos, probablemente ocasionados porque las cuerdas vibran a frecuencias distintas (patología referida a un aumento de masa de una de las cuerdas).
- ✓ Presencia de ruido sobre los 2000Hz, ocasionado por un cierre parcial (hiatus) de las cuerdas, que genera ruido por turbulencia.

El espectro que se muestra a continuación (figura 6.1) representa a dos señales, una normal (Roja) y la otra patológica (blanca), donde claramente se pueden visualizar irregularidades en relación a la claridad de la frecuencia fundamental y el valor de los armónicos.

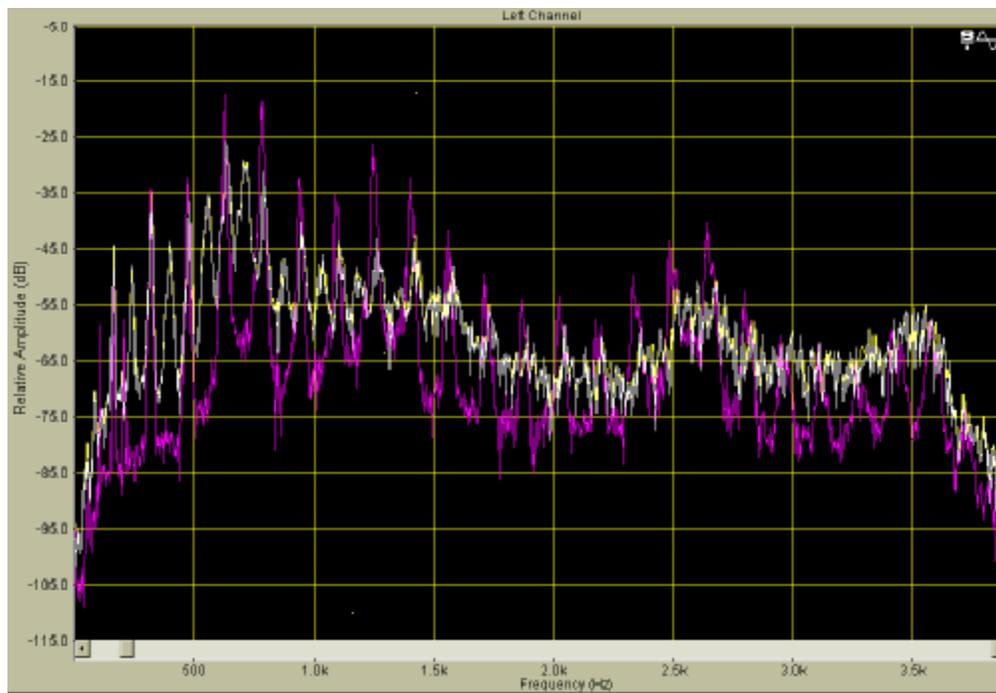


Fig. 6.1 – Espectrograma de un hablante masculino con voz normal y otro con voz patológica. (Fuente: Elaboración propia).

Ahora bien, gracias a la opinión de especialistas, que auditivamente emitían un diagnóstico con respecto a estas muestras podemos decir que realmente corresponden a pacientes o hablantes patológicos.

Es de total claridad visualizar que existe presencia de ruido por sobre los 2500 Hz., y por ende el valor de las formantes es sobre ésta frecuencia mucho más ilegible.

Sin embargo, se torna complicado diagnosticar, por medio de un análisis acústico algún tipo de patología, pero si se puede decir que en la mayoría de los caso la principal irregularidad se presentó a nivel de las cuerdas vocales.

6.2.3 Formantes

Para el análisis de las formantes se consideraron a tres hablantes masculinos y tres hablantes femeninos de la base de datos, donde a cada uno se les obtuvo el valor de las tres primeras formantes de cada vocal por las características propias que poseen. En las tablas 6.5 y 6.6 se muestran los valores formánticos de las vocales para hombres y mujeres obtenidos por medio del programa Praat.

	Fo (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/a/	137.6	706.8	1318.5	2360.0
/e/	135.6	482.4	2341.2	2477.8
/i/	135.1	355.7	2197.7	9584.8
/o/	133.1	478.7	874.8	2530.6
/u/	142.8	366.2	771.6	2465.1

Tabla 6.5 – Tabla representativa de los valores formánticos para los hablantes masculinos.
(Fuente: elaboración propia).

	Fo (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/a/	232.0	769.7	1420.3	2564.8
/e/	229.7	483.8	2147.5	2783.7
/i/	236.6	432.8	2353.3	2995.2
/o/	234.5	499.6	974.1	2845.4
/u/	230.1	456.0	879.4	2790.0

Tabla 6.6 – Tabla representativa de los valores formánticos para los hablantes femeninos.
(Fuente: elaboración propia).

A continuación, en las tablas 6.7 y 6.8 se muestran los valores de las formantes de un estudio realizado en Argentina, donde fueron tomadas un mayor número de muestras (40 Hombres y 40 mujeres).

	Fo	F1	F2	F3
/a/	127	830	1350	2450
/e/	125	430	2120	2628
/i/	130	290	2295	2915
/o/	124	510	860	2480
/u/	124	335	720	2380

Tabla 6.7 – Tabla representativa de los valores formánticos para los hablantes masculinos del estudio realizado en Argentina. (Fuente: M Guirao y AM Borzone de Manrique en el año 1975).

	Fo	F1	F2	F3
/a/	205	330	1553	2890
/e/	205	330	2500	3130
/i/	207	330	2765	3740
/o/	204	546	934	2966
/u/	204	382	740	2760

Tabla 6.8– Tabla representativa de los valores formánticos para los hablantes femeninos del estudio realizado en Argentina. (Fuente: M Guirao y AM Borzone de Manrique en el año 1975).

Si realizamos una comparación de ambos estudios es posible percibir una similitud en los valores de las frecuencias formánticas. En la siguiente tabla ejemplificaremos que porcentaje de variación existen entre ambos estudios considerando tanto a los hablante femeninos como a los masculinos.

hombres	mujeres	Total variación
10.5%	15.8%	13.2%

Tabla 6.9 – Tabla de variación porcentual de ambos estudios. (Fuente: Elaboración propia).

6.2.4 Carta de Formantes

Para el análisis de la voz normal se consideraron las cartas de las formantes, para visualizar mediante el cruce de F_1 y F_2 la distribución de las vocales, considerando el grado de elevación del dorso de la lengua y el grado de apertura del tracto vocal.

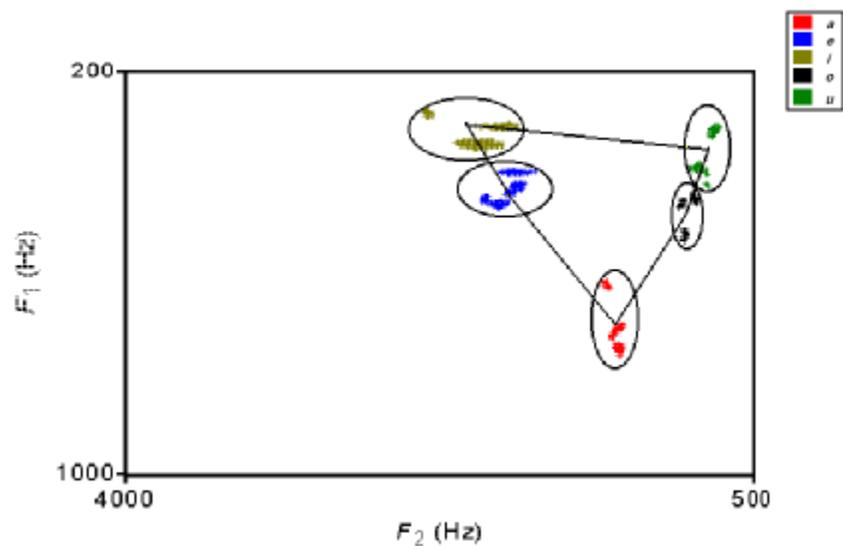


Fig 6.2 – Carta de Formantes de tres locutores Masculinos. (Fuente: Elaboración Propia)

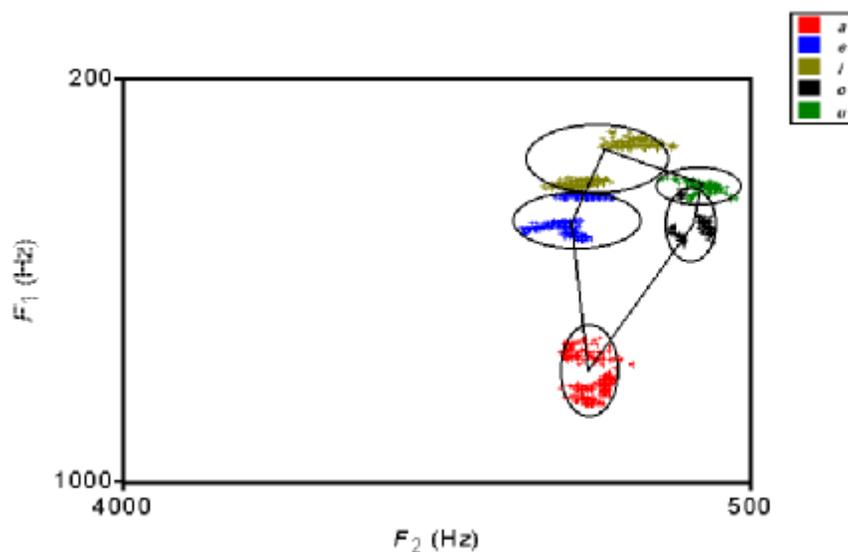


Fig 6.3 - Carta de Formantes de tres locutores Femeninos. (Fuente: Elaboración Propia)

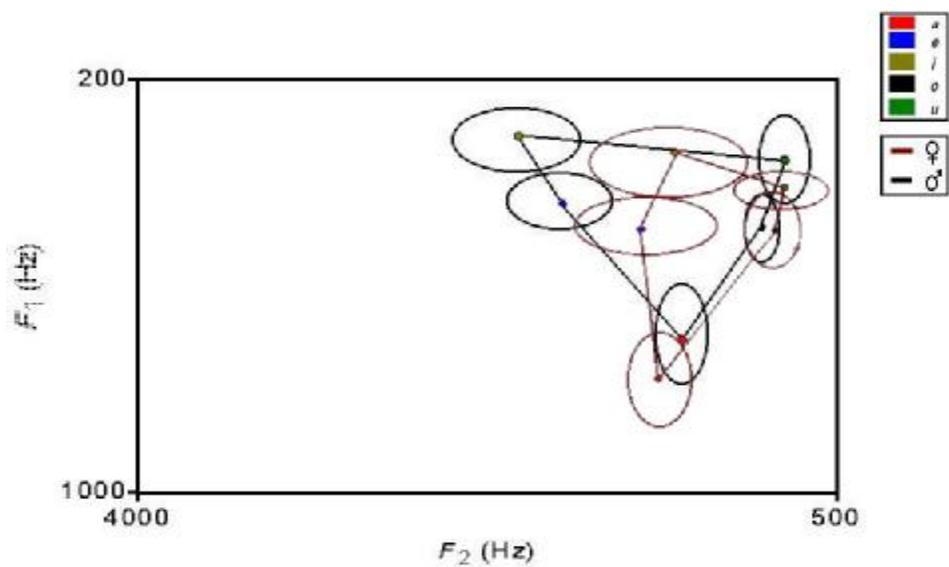


Fig 6.4 – Triángulo vocálico generalizado considerando hombres y mujeres realizados en la base de datos. (Fuente: Elaboración Propia)

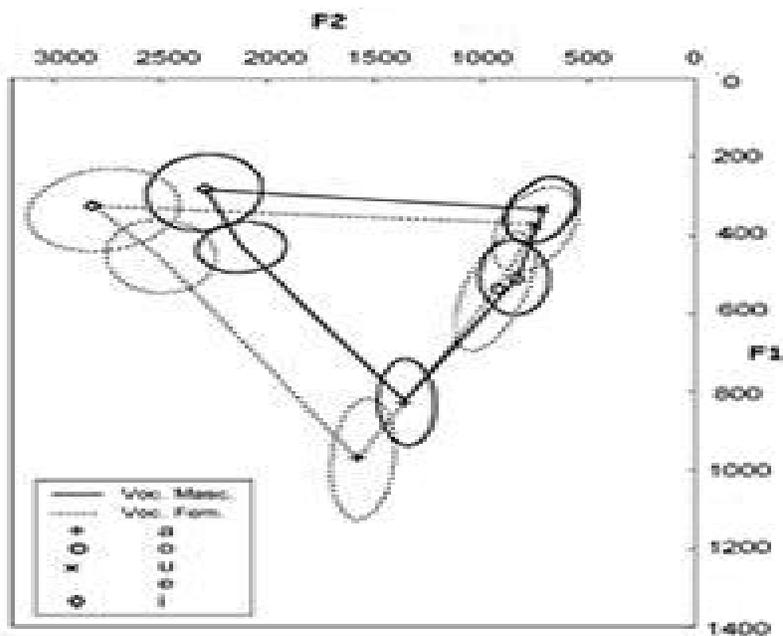


Fig 6.5 – Triángulo vocálico generalizado para hombres y mujeres según estudio rioplatense. (Fuente: M Guirao y AM Borzone de Manrique en el año 1975).

Si comparamos la Fig. 6.4 y 6.5 se puede decir que los triángulos vocálicos de los hombres (líneas oscuras) poseen una estructura similar considerando la abertura bucal y el largo de la cavidad oral. Por el contrario, en el caso de las mujeres, para el estudio realizado en Argentina existe una relación frecuencial superior producto de una mayor abertura de la cavidad oral, en comparación a la carta obtenida de la base de datos.

Por lo tanto, si se consideran las formantes en forma individual, se traduce que los valores frecuenciales de la segunda formante son mayores en el estudio realizados en Argentina que los evaluados en la base de datos, dando como resultado un estrechamiento en la conformación del triángulo vocálico para éste último.

6.3 Fonemas y coarticulación

Luego de seleccionar y analizar a los hablantes, tanto masculinos como femeninos, se pueden hacer las siguientes observaciones considerando los tres fonemas consonánticos elegidos:

6.3.1 Nasales

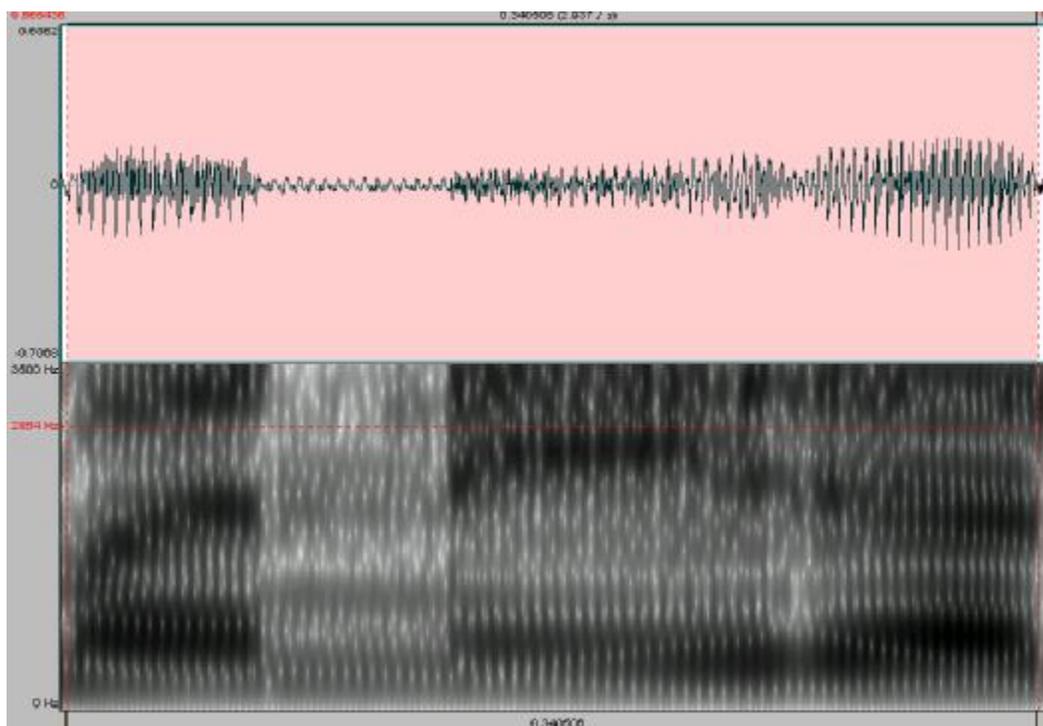


Fig. 6.6- Espectrograma de la palabra "MANERA". (Fuente: Elaboración propia).

En el caso del fonema /n/, las diferencias entre los hablantes masculinos son mínimas, es más, los valores de las formantes para cada uno son similares. Para el caso de los hablantes femeninos las frecuencias formánticas tienen un leve aumento, pero esto es producto de su conformación fisiológica.

Si se observa el espectrograma de la figura 6.6, es evidente que existe una variación de la energía para éste fonema (parte blanca) y las características de

sus formantes se ven alteradas considerando que se encuentra entre medio de dos fonemas vocálicos.

6.3.2 Lateral

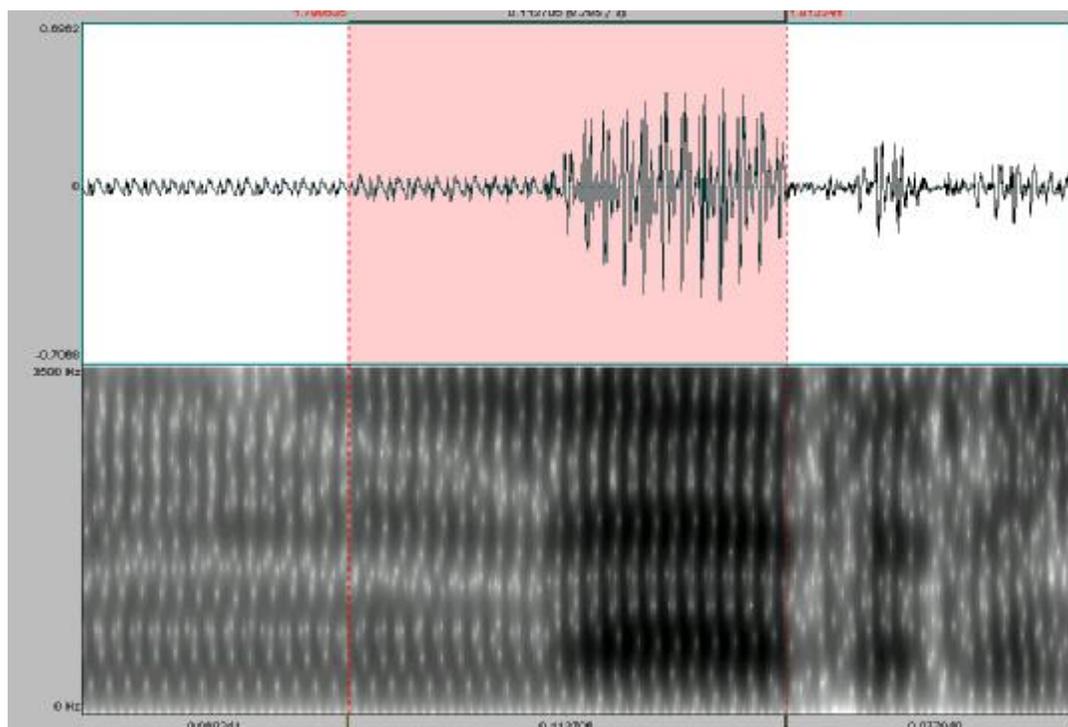


Fig. 6.7 – Espectrograma de la palabra “LA”. (Fuente: Elaboración propia).

Considerando ahora el fonema /l/, pero de diversas posiciones es posible darse cuenta que el porcentaje de error es mínimo para cada hablante, ya sean masculinos o femeninos.

En la figura 6.7 está representada la palabra “LA”, donde se puede visualizar claramente en el espectrograma la diferencia en las formantes de un fonema consonántico y un fonema vocálico. En otras palabras, la F_1 aparece a una frecuencia más baja que el correspondiente de la vocal con la que forma sílaba, y se ve poco influido por ella, es relativamente independiente de su vocal silábica; F_2 es también relativamente independiente del de la vocal independiente.

Para el caso de las otras dos posiciones de éste fonema se puede decir que los valores de las formantes varían, y tiene absoluta relación con el hecho de considerar que vocal la precede o la antecede.

6.3.3 Fricativa

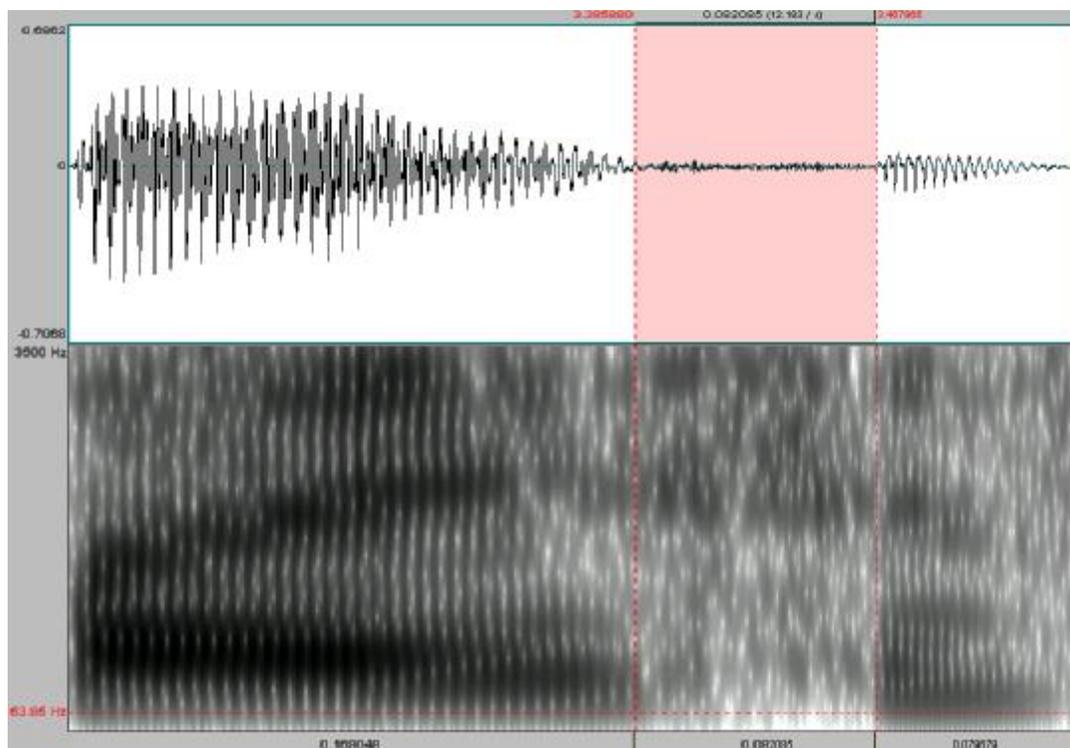


Fig. 6.8- Espectrograma de la palabra "PADECEN". (Fuente: Elaboración propia).

Para el caso de los fonemas fricativos, resulta difícil realizar un estudio exhaustivo por las características que éstas consonantes poseen. El principal impedimento enfrentado en esta investigación fue el lugar de concentración de la energía de las formantes, puesto que el sistema utilizado, como se explicó en el capítulo 5, es restringido y las formantes para éste fonema de resonancia alta están por sobre los 4000 Hz, por lo que nuestro sistema es incapaz de registrar dicha información, traduciéndose en que los datos obtenidos con la ayuda del

programa Praat son valores erróneos propios del programa y teóricamente imposibles.

6.3.4 Resultados obtenidos en medición de consonantes

A continuación se muestra en la tabla 6.10 los valores promedio de las formantes, correspondientes al fonema nasal /n/, fonema fricativo /s/ y el fonema consonántico /l/ considerando sus distintas posiciones:

	/n/	/l/-	-/l/-	-/l/	/s/
F1 (Hz)	425.6	446.5	377.75	455	1807.5
F2 (Hz)	1683.1	1558.2	1637.6	1461	2384.3
F3 (Hz)	2509.1	2765.5	2554.7	2480	3161.2

Tabla 6.10 - Frecuencias formánticas promedio de las personas analizadas. (Fuente: Elaboración propia).

Si estos resultados son comparados con otros estudios ya existentes, como por ejemplo los realizados por Quilis, se pueden validar los resultados obtenidos en éste trabajo ya que su estudio fue hecho para la lengua española.

Quilis	/n/	/l/-	-/l/-	-/l/
F1 (Hz)	361	327	328	343
F2 (Hz)	1400	1587	1534	1528
F3 (Hz)	2372	2603	2576	2575

Tabla 6.11 – Estudio realizado por Quilis para los fonemas /n/ y /l/ en diversas posiciones. (Fuente: Elaboración propia).

Este estudio realizado por Quilis (Tabla 6.11), considera a personas de ambos sexos y entrega un promedio para las frecuencias formánticas de cada fonema. Si consideramos el fonema /s/ en el estudio realizado por Quilis se puede determinar un rango frecuencial a partir de los 3500 Hz.

Basándonos en el estudio de Quilis para las frecuencias de las Formates de los Fonemas de la lengua española, podemos notar una diferencia en los valores de aproximadamente del 12% con respecto a los valores evaluados a partir de la Base de Datos.

6.4 Prosodia

Los resultados de las variaciones de la frecuencia Fundamental y las curvas melódicas estilizadas obtenidas para cada persona analizada, según su zona, son los siguientes:

Zona	Media(Hz)	Desviación Estándar (Hz)	Rango de Frecuencia (Hz)
Norte	166	35	122
Centro	114	11.1	32
Sur	160	18	52

Tabla 6.12 – Valores estadísticos de las variaciones de la frecuencia fundamental para personas de distintas zonas del país. (Fuente: Elaboración propia).

Zona Norte

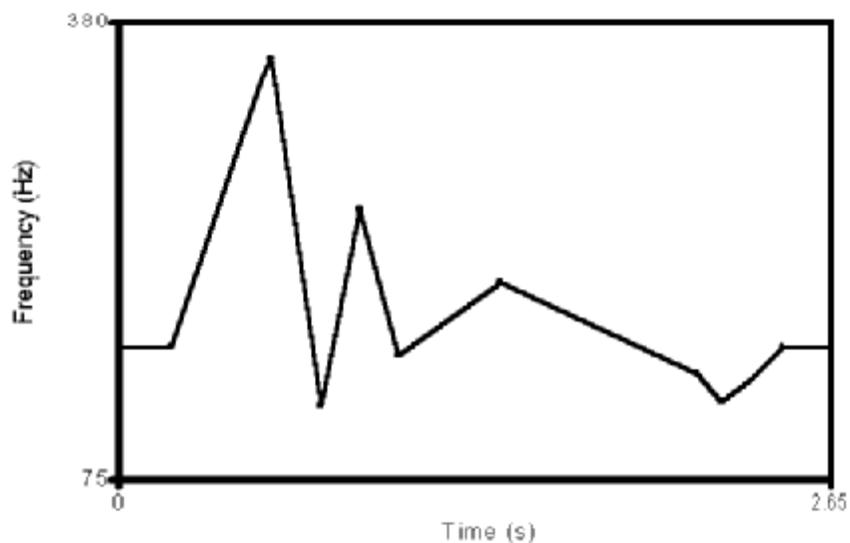


Fig. 6.9- Curva melódica estilizada de la frase “El Dieciocho de Septiembre se celebra” para el locutor 129 de la zona Norte. (Fuente: elaboración propia).

Zona Centro



Fig. 6.10- Curva melódica estilizada de la frase “El Dieciocho de Septiembre se celebra” para el locutor 287 de la zona Centro. (Fuente: elaboración propia).

Zona Sur

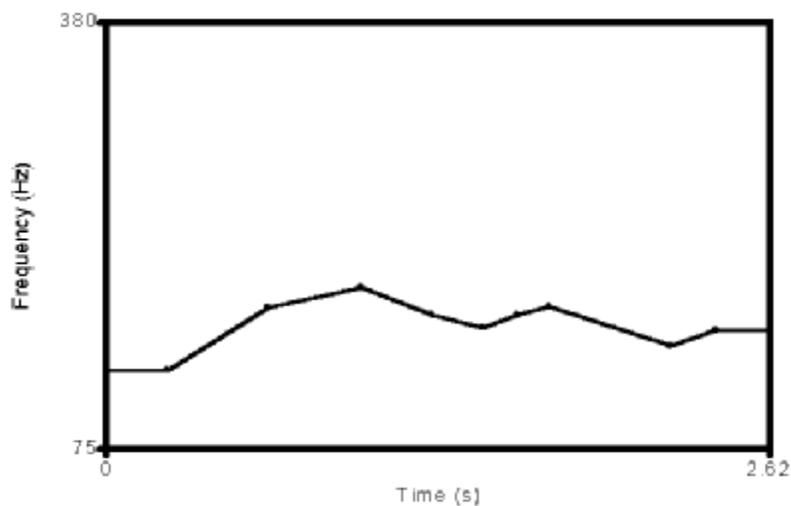


Fig. 6.11- Curva melódica estilizada de la frase “El Dieciocho de Septiembre se celebra” para el locutor 69 de la zona Sur. (Fuente: elaboración propia).

Si observamos la estructura de cada curva melódica estilizada para cada persona, se aprecia que la persona de la zona norte tiene tendencia a variar su frecuencia fundamental (de mayor manera) con respecto a las otras dos personas además de tener una frecuencia Fundamental promedio mayor a las personas de la zona Centro y Sur que poseen una rango de variación de F_0 más restringido.

Ahora bien si consideramos, los patrones locales, y más bien los movimientos finales, es decir, ascendentes y descendentes, para los tres casos ocurre un comportamiento similar.

Si se consideraran los mismos intervalos de tiempo para las tres zonas, se puede asumir una tendencia generalizada, es decir con similares puntos de inflexión y comportamientos ascendentes y descendentes de F_0 .

7 CONCLUSIONES

A continuación se presentan las conclusiones que se generan con el desarrollo íntegro, en función de los análisis de resultados obtenidos en el capítulo anterior.

7.1 Efectividad del sistema electroacústico

El hecho de ocupar un híbrido telefónico para realizar las grabaciones de los hablantes para la Base de Datos restringe el ancho de banda de la señal a grabar. La existencia o mejor dicho la aplicación de otros sistema de grabación hubiesen permitido realizar otro tipos de análisis, donde se puedan considerar frecuencias mayores a los 3500 Hz.

Una ventaja muy importante de éste sistema electroacústico tiene relación con el bajo costo que significó su utilización y el fácil manejo para operarlo.

7.2 Formato de la base de datos

Considerando Bases de Datos de locutores construidas en otros países se llegó a la determinación de ciertos parámetros al momento de realizar las grabaciones. Éstos parámetros, tales como frecuencia de muestreo, cuantización y sistema electroacústico en sí (Interfaz telefónica), fueron analizados en detalle para garantizar un óptimo desempeño.

La idea de ocupar una frecuencia de muestreo finalmente de 8 KHz. y una cuantización de 16 bits fue exclusivamente por concepto de espacio ya que no se ve alterada la muestra por el ancho de banda que posee la voz y la respuesta de frecuencia que el sistema tenía.

7.3 Comparación Micrófono –Teléfono

Podemos concluir que el 6% de diferencia existente en ambos sistemas de grabación es debido a la restricción del ancho de banda que posee la señal telefónica, si bien, existen parámetros (como es el caso de los descriptores acústicos Shimmer y NHR) que no son fiables de ser evaluados por medio de esta Base de Datos, los demás parámetros de interés no sufren una variación significativa con respecto a los valores evaluados fielmente por el micrófono.

7.4 Metodologías

Quizás uno de los temas más relevantes en nuestro trabajo de investigación, fue la construcción de metodologías para las distintas áreas a la que estábamos enfocados.

La utilización de programas idóneos, la colaboración de especialistas y la disposición de entregar, o más bien, de demostrar hacia donde pueden estar enfocadas trabajos posteriores concernientes a temas como Voz Normal, Voz Patológica, y Lingüística.

7.5 Análisis de los resultados

7.5.1 Tono Medio Hablado

En la evaluación del tono medio hablado es posible hacer varios alcances.

- ✓ La utilización de una serie de números hablada en forma continua entrega una representación confiable del valor de la frecuencia fundamental de un hablante, ya que no se utilizan los elementos suprasegmentales (acento, entonación, melodía, etc...) que producen una variación natural del pitch.

- ✓ En algunos casos, la utilización de los días de la semana o los meses del año pueden entregarnos también un valor frecuencial representativo.
- ✓ Existe una tendencia a variar ésta frecuencia al momento de terminar la serie automática.
- ✓ La frecuencias fundamentales analizadas coinciden con el estudio realizado por Quilis el año 1991 tanto para locutores masculinos como para locutores femeninos.

7.5.2 Voz Normal y Patológica

- ✓ No es posible detectar patologías del habla a nivel resonancial, debido a los cambios en la amplitud de la señal a grabar, que son producidas por la grabación telefónica.
- ✓ Las alteraciones que en algunos casos sufren los descriptores acústicos de la voz (Jitter y Shimmer), son ocasionados por el sistema de grabación de muestras.
- ✓ Si bien los valores de las frecuencias formánticas de los fonemas vocálicos coinciden con los valores obtenidos por otros estudios y son considerados normales, el análisis de voz normal y voz patológica no es totalmente factible debido a las alteraciones de sufren el Jitter y Shimmer.
- ✓ Debido a la respuesta de frecuencia del sistema original no es posible realizar, en forma confiable, estudios relacionados con voz patológica.

7.5.3 Formantes

- ✓ Tomando en cuenta el análisis realizado en el capítulo 6 referente a las formantes, que tiene relación con nuestros análisis y los realizados en Argentina, se puede decir que :
 - En primer lugar, el valor de la frecuencia fundamental para los hablantes argentinos es menor que para los realizados con las muestras de la base de datos con locutores chilenos.
 - El porcentaje de variación, tomando en cuenta ambos estudios es de alrededor del 13 %, lo que indica que existe una similitud en los datos, a pesar de que los sistemas electroacústicos implementados fueron diferentes.

7.5.4 Carta de Formantes

El análisis acústico de las vocales ofrece más ventajas que un análisis realizado desde el punto de vista articulatorio.

- ✓ En primer lugar, el número de parámetros que es necesario tener en cuenta en el nivel articulatorio para especificar con exactitud la posición de los órganos fonadores en la emisión de una vocal es muy elevado, mientras que los parámetros acústicos necesarios son mucho más reducidos.
- ✓ Considerando el estudio hecho en Argentina y los análisis realizados, se puede concluir que en los locutores chilenos el rango del movimiento de la lengua (representado por la F_2) y el rango de apertura de la boca (representada por F_1) es menor con respecto al locutor argentino.
- ✓ La posición de la lengua en las vocales /e/ e /i/, en los locutores femeninos chilenos se centraliza, con respecto al estudio realizado en Argentina (M. Guirao y AM Borzone de Manrique, 1975).

7.5.5 Fonemas

- ✓ El análisis espectral de los fonemas con frecuencias formánticas sobre los 4000 Hz. se dificulta por la respuesta de frecuencia propia del sistema. Es el caso de los fonemas fricativos y otros donde la mayor concentración de energía está por sobre éste rango.
- ✓ Para los fonemas consonánticos áfonos, es difícil llegar a realizar un análisis, ya que al no tener participación las cuerdas vocales, las formantes se encuentran en el ancho de banda fuera de nuestro sistema de grabación original.

7.5.6 Prosodia

- ✓ Considerando las tres zonas analizadas, es posible darse cuenta que existe una tendencia, para los tres casos, de conservar un patrón melódico final. Esto se observa en el comportamiento del último tramo de la curva estilizada para cada locutor, que tiende a tomar un valor constante.
- ✓ De los resultados obtenidos en las variaciones de la frecuencia fundamental se puede desglosar que en la zona norte existe una variación de frecuencia mayor con respecto a las otras dos zonas (ejemplificado en la tabla 6.8) debido a los valores de la desviación estándar y el rango de frecuencia en que oscila el tono fundamental.
- ✓ Es de suma importancia destacar que los análisis realizados corresponden sólo a tres locutores del territorio chileno, por lo tanto es necesario realizar un análisis con la totalidad de muestras de la base de datos para obtener un acercamiento al patrón melódico chileno.

7.6 Alcances de la Base de Datos Fonética

Las futuras aplicaciones de la base de datos fonética son las siguientes:

- ✓ Obtención de un valor representativo normal del Tono Medio Hablado y de las tres primeras formantes para el hablante chileno masculino y femenino, cuyos estándares pueden tener una utilización en la evaluación clínica.
- ✓ Determinación de la carta de formantes para el hablante chileno masculino y femenino, con el fin de realizar una descripción acústico-articulatoria de los sonidos vocálicos y comparar los resultados entre las distintas zonas geográficas del país además de otras lenguas españolas.
- ✓ Caracterización acústica de los fonemas del español de Chile mediante un análisis de su estructura formántica.
- ✓ Determinación de la entonación de la voz por medio de las frases y su acentuación por medio de las palabras, con el fin de encontrar curvas melódicas asociadas a las distintas zonas de nuestro país.

Se propone la creación de una nueva base de datos fonética con fines de estudio de voces patológicas (condiciones acústicas controladas, mejorar la respuesta del sistema electroacústico), debido a los resultados obtenidos con el análisis de las muestras registradas.

8 REFERENCIAS

- [1] Bernal, J., Bobadilla, J. Y Gómez, P. ;Reconocimiento de Voz y Fonética acústica; México: Alfa Omega; 2000.
- [2] Bernal, J., Bobadilla, J. Y Gómez, P. ;Estudios de Fonética Experimental; Barcelona :PPU. 1999.
- [3] Communication technology Group, University of Zaragoza, España, 1997.
- [4] Estruch; M.; Garrido; J, Listerri; J.; Aproximación fonética al estudio de la entonación. Trabajos de fonética (Paper); España; 1996.
- [5] Estruch; M.; Garrido; J, Listerri; J.; Técnicas y procedimientos para la representación de las curvas melódicas (Paper); Universidad Autónoma de Barcelona; España; 2003.
- [6] Garcia, Rafael; Cobeta, Ignacio; Diagnostico y tratamientos de los trastornos de la voz; Ed. García; Santiago; Chile; 1991.
- [7] Garrido; J, Listerri; J.; Estudio comparado de las características prosódicas del español en dos modalidades de lectura (Paper); Trabajos de fonética experimental; España; 1995.
- [8] Gil, J.; Listerri; J.; Fonética y Fonología del español en España (Paper); Universidad Autónoma de Barcelona; España; 2003.
- [9] González Rodríguez, Joaquín; Ortega García, Javier; Procesado de Voz: Reconocimiento de mensaje y locutor; EUIT de telecomunicación; Madrid; España; 1997.
- [10] González Rodríguez, Joaquín; Influencia y Compensación del entorno Acústico en Sistemas de Reconocimiento Automático de Locutores; Director: Hernández; Luis; Madrid; España; 1999.
- [11] International Phonetics Association Internet:
<http://www.unil.ch/ling/phonetique>.
- [12] Jackson, Menaldi. ; La Voz Normal. Ed. Panamericana; Buenos Aires; 1992.

- [13] Jackson, Menaldi. ; La Voz Patológica Ed. Panamericana; Buenos Aires 2002.
- [14] Kent, R. , Read, Ch. Acoustic Analysis of Speech; Canada; Delmar. 2002.
- [15] Ladefoged, Peter; Elements of Acoustic Phonetics; Chicago; United States of America; 1962.
- [16] Listerri; J.; sitio web: <http://www.liceau.uab.es/>.
- [17] Malcolm; J.; Crocker, I.; Encyclopedia of Acoustics (Volumen IV); United States of America; 1997.
- [18] Miyara, Federico; Acústica del Tracto Vocal y Voz Humana
- [19] Pardo, J. M.; Sistema de producción del habla, Apuntes de Ingeniería Neusensorial; Universidad Politécnica de Madrid; España; 2002
- [20] Quilis, Antonio; Fonética Acústica de la Lengua Española; Madrid; España; 1988.
- [21] Quilis, Antonio; Fonética General de la Lengua Española; Madrid; España; 1991.
- [22] Roca, Ricardo; sitio web: <http://www.ricardoroca.cl/>.
- [23] Zañartu, Matías; Barrera, Andrés.; Espinoza, Victor; Painepan, Juan Carlos; Trujillo, Rodrigo; Guías de Laboratorio de Física Acústica. Escuela de Fonoaudiología; Universidad Mayor; Santiago; Chile 2003.
- [24] Zañartu, Matías; Apuntes Cátedra Física-Acústica; Escuela de Fonoaudiología; Universidad Mayor; Santiago; Chile 2003.

9. ANEXOS

9.1 Espectrogramas de dos hablantes masculinos y dos hablantes femeninos correspondientes a voces patológicas.

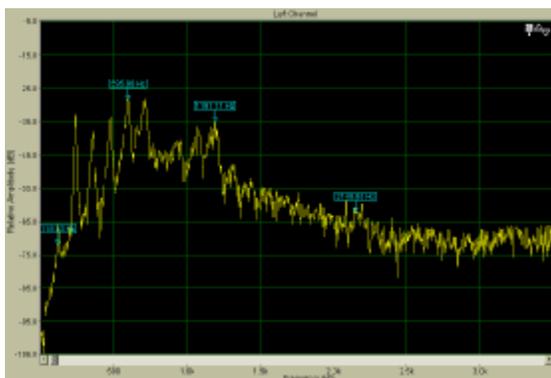


Fig. 9.11 – Espectrograma del locutor 118 hablante masculino. (Fuente: Elaboración propia).

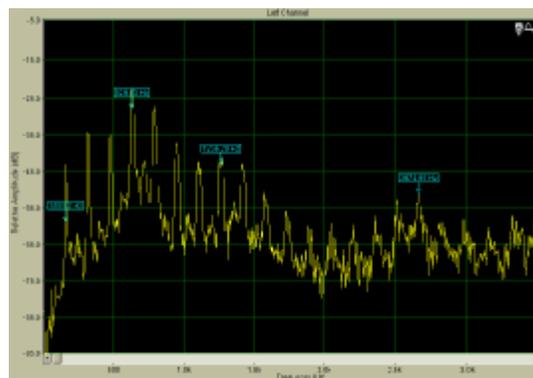


Fig. 9.12 – Espectrograma del locutor 129 hablante masculino. (Fuente: Elaboración propia).

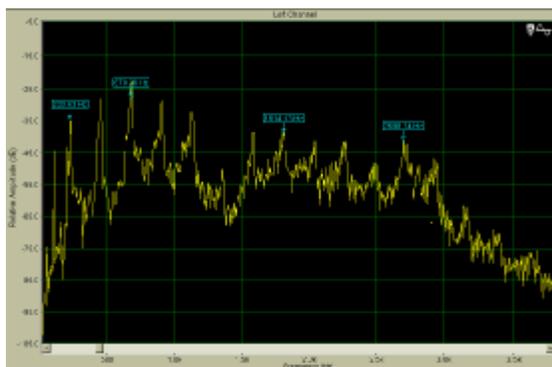


Fig. 9.13 – Espectrograma del locutor 12 hablante femenino. (Fuente: Elaboración propia).

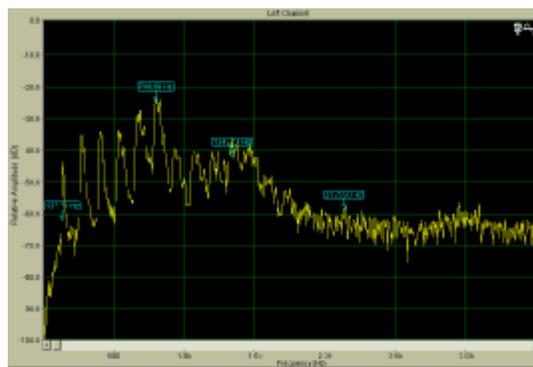


Fig. 9.14 – Espectrograma del locutor 97 hablante femenino. (Fuente: Elaboración propia).

9.2 Espectrogramas de dos hablantes masculinos y dos hablantes femeninos correspondientes a voz normal.

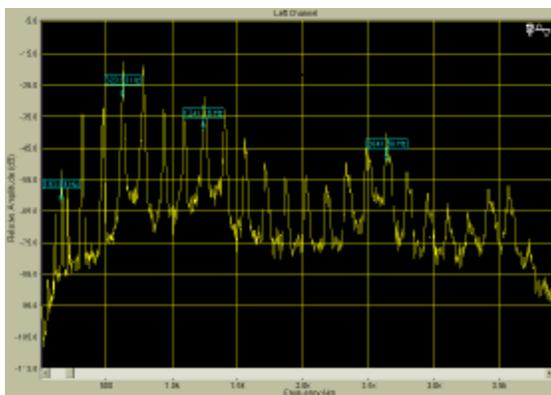


Fig. 9.21 – Espectrograma del locutor 11 hablante masculino. (Fuente: Elaboración propia).

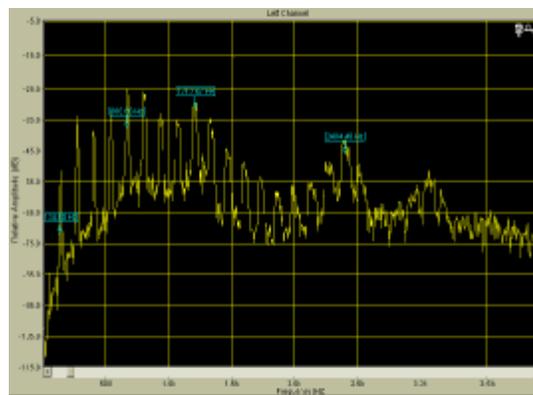


Fig. 9.22 – Espectrograma del locutor 56 hablante masculino. (Fuente: Elaboración propia).

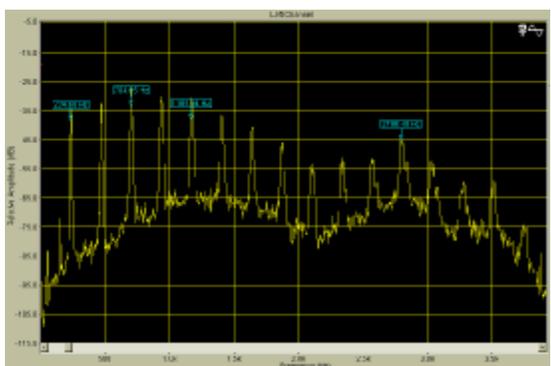


Fig. 9.23 – Espectrograma del locutor 58 hablante femenino. (Fuente: Elaboración propia).

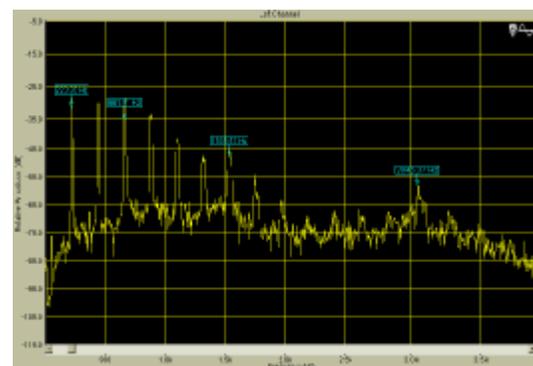


Fig. 9.24 – Espectrograma del locutor 243 hablante femenino. (Fuente: Elaboración propia).

9.3 Cartas de Formantes individuales para personas con voz normal.



Fig. 9.31 – Carta de formante locutor 56 hablante masculino. (Fuente: Elaboración propia).



Fig. 9.32 - Carta de formante locutor 11 hablante masculino. (Fuente: Elaboración propia).



Fig. 9.33 - Carta de formante locutor 199 hablante masculino. (Fuente: Elaboración propia).



Fig. 9.34 - Carta de formante locutor 58 hablante femenino. (Fuente: Elaboración propia).

9.4 Tablas de formantes y descriptores acústicos para los hablantes de voz normal y voz patológica.

Persona		Fo (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)	Shim(%)	Jitt(%)	NHR(%)
11	/a/	155.3	751.2	1366.4	2513.1	2.8	0.30	25.6
	/e/	149.7	499.0	2998.1	2613.0	3.1	0.20	21.5
	/i/	148.0	352.1	2271.0	2568.3	3.3	0.20	20.0
	/o/	147.1	533.0	890.1	2703.0	1.7	0.30	28.5
	/u/	167.1	398.1	821.3	2603.1	1.8	0.45	24.9
56	/a/	130.0	747.0	1241.0	2134.0	3.1	0.48	20.7
	/e/	126.0	518.0	1932.0	2512.0	2.9	0.46	19.8
	/i/	121.0	403.0	2301.0	2482.0	3.3	0.34	19.9
	/o/	121.0	430.0	823.0	2402.0	2.5	0.20	27.3
	/u/	124.0	301.0	692.0	2304.0	1.2	0.20	30.4
199	/a/	127.6	622.1	1348.1	2433.0	1.3	0.40	24.7
	/e/	131.2	430.1	2093.5	2308.4	1.8	0.20	23.1
	/i/	136.3	312.0	2021.0	23704.0	2.2	0.27	22.7
	/o/	131.1	473.2	911.3	2486.7	1.5	0.27	23.3
	/u/	137.3	399.4	801.5	2488.2	1.2	0.31	29.6
58	/a/	232.2	840.1	1382.2	2387.0	3.1	0.28	21.2
	/e/	230.7	485.0	1745.6	2517.0	5.5	0.26	22.0
	/i/	232.0	415.2	1865.0	2776.3	4.8	0.50	20.1
	/o/	228.5	526.1	886.1	2878.1	1.8	0.20	30.0
	/u/	232.0	442.1	864.5	2603.0	1.6	0.10	28.7
243	/a/	221.7	727.1	1276.7	2506.3	3.8	0.41	22.2
	/e/	215.3	437.1	2306.9	2734.0	2.5	0.22	20.9
	/i/	223.7	424.1	2503.0	3050.0	1.8	0.20	27.5
	/o/	220.1	460.5	939.1	2857.0	2.2	0.31	31.8
	/u/	216.9	461.0	854.0	2963.1	1.3	0.23	31.0
308	/a/	242.0	742.0	1602.0	2801.0	2.6	0.14	23.4
	/e/	243.0	529.2	2390.0	3100.0	2.1	0.31	26.1
	/i/	254.0	459.2	2692.0	3159.2	3.5	0.40	25.2
	/o/	255.0	512.3	1097.0	2801.0	1.3	0.19	32.7
	/u/	241.3	465.0	919.8	2804.0	1.4	0.38	27.8

Tabla 9.41 – Tabla comparativa de las frecuencias formánticas de las vocales de personas con voz normal y sus respectivos descriptores. (Fuente: Elaboración propia).

Persona		Fo	F1	F2	F3	Shim(%)	Jitt(%)	NHR(%)
118	/a/	116.2	636.0	1131.0	1990.0	12.7	0.70	16.9
	/e/	115.4	439.0	1780.0	2323.0	6.0	0.40	19.6
	/i/	114.0	316.0	1997.0	2573.0	3.4	0.30	23.4
	/o/	114.0	517.0	831.0	1993.0	5.6	0.60	21.1
	/u/	110.0	405.0	663.0	2131.0	5.7	1.00	17.1
129	/a/	158.1	685.0	1243.0	2501.0	5.4	0.60	12.4
	/e/	158.0	458.0	2017.0	2789.0	11.1	2.20	4.8
	/i/	162.0	403.0	2102.0	2600.4	11.5	0.39	10.1
	/o/	157.0	517.9	982.0	2691.0	13.3	0.89	19.2
	/u/	149.0	439.0	738.1	2493.0	8.8	2.21	18.0
222	/a/	102.3	771.0	1189.0	2500.0	8.5	0.50	16.3
	/e/	105.0	469.0	1851.0	2308.0	7.3	0.56	19.6
	/i/	106.4	287.0	2087.0	2519.0	3.9	0.50	18.4
	/o/	103.0	524.0	828.1	2385.0	7.3	0.40	17.2
	/u/	100.1	386.0	870.0	2544.0	5.0	0.90	19.1
12	/a/	218.0	681.0	1571.1	2567.0	9.8	0.60	14.3
	/e/	213.0	467.0	1063.1	2548.0	7.2	0.50	16.2
	/i/	220.3	443.0	2230.0	2787.0	3.5	0.40	15.6
	/o/	208.0	540.0	829.0	2497.0	6.2	0.38	19.1
	/u/	205.9	445.0	770.9	2354.0	5.9	0.46	21.3
97	/a/	136.0	769.0	1393.0	2149.0	5.4	0.50	18.7
	/e/	123.4	493.0	2029.0	2603.0	2.4	0.30	19.2
	/i/	127.1	381.0	1359.0	2639.0	4.5	0.80	19.2
	/o/	129.0	513.0	859.0	2489.0	4.2	0.80	20.1
	/u/	141.0	428.0	1013.0	2581.0	3.1	0.70	19.0
189	/a/	202.0	879.0	1525.0	2329.0	5.0	0.40	17.2
	/e/	194.0	448.0	812.0	2678.0	5.8	2.30	12.9
	/i/	206.0	403.0	2390.0	2900.0	4.6	0.90	16.5
	/o/	200.0	476.0	894.0	2579.0	4.8	1.30	18.0
	/u/	201.0	402.0	749.0	2379.0	4.2	0.70	16.1

Tabla 9.42 – Tabla comparativa de las frecuencias formánticas de las vocales de personas con voz patológica y sus respectivos descriptores. (Fuente: Elaboración propia).

Hombres		/n/	/l/	-/l/	-/l/	/s/
persona 9	F1	446	396	333	416	1711
	F2	1419	1304	1504	1219	2445
	F3	2290	2548	2152	2126	3304
persona 56	F1	330	376	331	381	1688
	F2	1498	1401	1419	1334	2466
	F3	2532	2450	2201	2316	3232
persona 95	F1	377	380	338	433	1698
	F2	1530	1345	1406	1327	2357
	F3	2329	2460	2380	2280	3146
persona 105	F1	357	382	322	408	1803
	F2	1528	1438	1467	1301	2513
	F3	2432	2444	2357	2554	2960

Tabla 9.43 – Tabla de los valores formánticos para hablantes masculinos de las consonantes seleccionadas. (Fuente: Elaboración propia).

Mujeres		/n/	/l/	-/l/	-/l/	/s/
persona 34	F1	517	590	423	503	1890
	F2	1863	1843	1670	1609	2228
	F3	2358	3143	2664	2492	3162
persona 120	F1	478	520	411	497	no valor
	F2	1786	1786	1797	1442	no valor
	F3	2651	3127	3040	2789	no valor
persona 276	F1	444	468	461	531	no valor
	F2	1868	1719	1817	1317	no valor
	F3	2681	2997	2820	2245	no valor
persona 290	F1	456	460	403	468	no valor
	F2	1973	1630	2021	2137	2419
	F3	2800	2955	2824	3037	no valor

Tabla 9.44 – Tabla de los valores formánticos para hablantes femeninos de las consonantes seleccionadas. (Fuente: Elaboración propia).

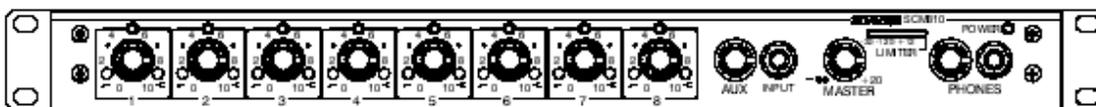
Sistema de Interfase AC-10H Clear-Com:



ESPECIFICACIONES:

Requerimientos de Potencia	18 mA
Respuesta de frecuencia	200 Hz - 12 kHz, +3 dB
Impedancia de Interfase	
Modo 2-Wire, Normal:	1,100 Ω
Modo 2-Wire, Low-Z:	4 Ω
Modo 3- or 4- Wire ,	
Transmisión de Salida:	200 Ω
Entrada Recibida:	500 Ω
Transmisión de Salida:	
Modo 2-Wire, Normal:	+8 dBm max into 200 Ω
Modo 2-Wire, Low-Z:	125 mV max into 4 Ω
Modo 3- or 4-Wire:	+4 dBm max into 600 Ω
Conectores de Entrada y Salida:	
(4) 5-way binding posts para interfase con otros sistemas	
(1) 3-pin conector XLR hembra para interfases Clear-Com	

Mezcladora Shure SCM810:



ESPECIFICACIONES:

Respuesta de frecuencia (Ref 1 kHz):	50 Hz a 20 kHz ± 2 dB; -3 dB de atenuación a 25 Hz
Distorsión armónica total:	<0,1% al nivel de salida de +18 dBV, 50 Hz a 20 kHz
Ruido de entrada equivalente:	-125 dBV (fuente 150)
Zumbido y ruido de entrada equivalente:	-123 dBV (fuente 150 Ω)
Zumbido y ruido de salida:	-90 dBV
LED de entrada:	Verde: canal activado, Rojo : -6 dB por debajo del nivel de "clipping"
Alimentación Phantom:	Circuito abierto de 48 VCC (6,8 k Ω).

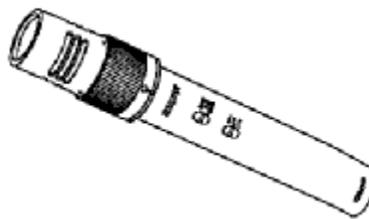
Grabador de Audio Digital ADAT-XT20™ 20 Bit:



ESPECIFICACIONES:

Formato de Grabación:	ADAT Tipo I (16-bit), ADAT Tipo II (20-bit) Grabación digital con cabezal rotatorio.
Canales de Audio:	8 Canales
Conversión Digital:	<i>Record (A/D):</i> 20 bit lineal, 128 veces de sobre muestreo, convertidor individual por canal; <i>Playback (D/A):</i> 20 bit lineal, 128 veces de sobre muestreo, convertidor individual por canal.
Razón de Muestreo:	44.1 /48kHz, Opcional.
Respuesta de Frecuencia:	20Hz-20 kHz, $\pm 0.5\text{dB}$
Rango Dinámico:	102 dB, A-weighted
Distorsión:	0.003% THD
Wow and Flutter:	Inmensurable
Entradas/Salidas Análogas:	<i>(Balanceada)</i> 1 conector ELCO" (in/out), <i>(Desbalanceada)</i> 16 phono jacks (RCA) ,(8 entradas, 8 salidas)
Entradas/Salidas Digitales:	8-canales jacks Ópticos ADAT (1 entrada, 1 salida).

Micrófono de condensador Shure KSM141:



ESPECIFICACIONES:

Tipo de cápsula	Condensador de polarización permanente.
Respuesta de frecuencia	20–20.000 Hz
Patrón polar de captación	Cardioide/omnidireccional
Impedancia de salida	150 Ω real.
Interruptor de atenuación	0, 15 ó 25 dB de atenuación
Interruptor de respuesta de frecuencias bajas	Respuesta uniforme, –6 dB/octava a menos de 115 Hz, –18 dB/octava a menos de 80 Hz.
Alimentación Phantom	48 \pm 4 VCC
Rechazo en modo común	50 dB, 20 Hz hasta 20 kHz
Sensibilidad (1kHz;1 Pa = 94 dB SPL)	–37 dBV/Pa
Nivel de presión acústica (SPL)máx.	
5000 Ω de carga	145 (160, 170) dB
2500 Ω de carga	139 (154, 164) dB
1000 Ω de carga	134 (149, 159) dB
Rango dinámico	
5000 Ω de carga	131 dB
2500 Ω de carga	125 dB
1000 Ω de carga	120 dB
Relación de señal a ruido	80dB

