INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

MODELO MECÁNICO ACÚSTICO DEL OÍDO INTERNO EN RECONOCIMIENTO DE VOZ

TESIS

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:

M. en C. MARIO JIMÉNEZ HERNÁNDEZ

DIRECTORES DE TESIS:

DR. JOSÉ LUIS OROPEZA RODRÍGUEZ

DR. SERGIO SUÁREZ GUERRA

MÉXICO, D.F. 2013

SIP-14 BIS

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 11:00 horas del día 13 del mes de Mayo de 2013 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"MODELO MECÁNICO ACÚSTICO DEL OÍDO INTERNO EN RECONOCIMIENTO DE VOZ"

Presentada por el alumno:

JIMÉNEZ	**HERNÁNDEZ**	**MARIO**
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

B	0	9	1	6	9	3

aspirante de: **DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis

Dr. José Luis Oropeza Rodríguez

Dr. Sergio Suárez Guerra

Dr. Oleksiy Pogrebnyak

Dr. Ricardo Barrón Fernández

Dr. Héctor Manuel Pérez Meana

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN
DIRECCIÓN

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México D.F. el día 13 del mes mayo del año 2013, el (la) que suscribe Mario Jiménez Hernández alumno (a) del Programa de Doctorado en Ciencias de la Computación con número de registro B091693, adscrito al Laboratorio de Procesamiento Digital de Señales, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Sergio Suárez Guerra y Dr. José Luis Oropeza Rodríguez y cede los derechos del trabajo intitulado Modelo mecánico acústico del oído interno en reconocimiento de voz, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección mjimenezh@ipn.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

MJiménezH

Mario Jiménez Hernández

Nombre y firma

Resumen

Este trabajo de Tesis presenta un nuevo modelo mecánico acústico del oído interno basado en su respuesta física usando análisis por resonancia. Se fundamenta en la mećanica de fluidos para describir el comportamiento de la perilinfa y la endolinfa dentro de la escala media, la escala vestibular y la escala timpánica. Por ser fluidos incompresibles se utiliza la conservación del momentum para describir su movimiento y el teorema de la divergencia para obtener sus condiciones límites para pequeñas amplitudes ignorando los términos no lineales. El comportamiento mecánico de la membrana basilar se analiza como un sistema de osciladores armónicos forzados amortiguados concatenados a partir del modelo propuesto por Lesser y Berkeley, considerando que la ecuación de onda que describe el movimiento de la membrana basilar es la condición límite de la diferencia de presiones sobre cada punto de la membrana y que las características físicas de masa, constante de elasticidad y resistencia mecánica a lo largo de la membrana presentan diferentes valores en función de la distancia desde el ápice hasta el helicotrema.

El nuevo modelo de análisis por resonancia considera que el sistema es excitado por una fuerza externa periódica y compleja producida por las vibraciones transmitidas al interior de la cóclea por la ventana oval, siendo esta fuerza la solución de la ecuación del oscilador forzado amortiguado. Se considera que el desplazamiento y la amplitud también son complejos, definiendo la impedancia mecánica compleja de entrada del sistema como la suma de la parte real dada por la resistencia mecánica y una parte imaginaria dada por la reactancia mecánica, expresando en forma polar los términos de las componentes de magnitud y fase. La identidad de Euler se usa para separar la parte real de la imaginaria y obtener una expresión para la amplitud del sistema que dependa sólo de las características físicas de masa, constante de elasticidad y resistencia mecánica a lo largo de la membrana basilar y de la frecuencia de la fuerza excitadora.

Con el nuevo modelo desarrollado se tiene la ventaja respecto a los modelos existentes de determinar la distancia a lo largo de la membrana basilar donde se presénta la amplitud máxima para todo el intervalo de frecuencias de la audición humana. Para su evaluación se hace la comparación con los modelos de la cóclea desarrollados por Peterson y Bogert (Método de integración numérica), de Allen (Análisis mediante la función de Green), de Neely (Método de diferencias finitas) y con las mediciones experimentales de Békésy, usando en cada experimento los parámetros físicos de la membrana basilar y las frecuencias de evaluación de cada modelo. Por último se muestra una aplicación para el reconocimiento de voz desarrollando una nueva forma de parametrización y su implementación en HTK, se propone un arreglo de filtros usando el método de diferencias finitas y el análisis por resonancia para determinar el intervalo de frecuencias de cada filtro, obteniendo en todos los experimentos resultados satisfactorios.

Abstract

This Thesis presents a new acoustic mechanical model of the inner ear based in their physics response using resonance analysis. It is based in the fluid mechanics for described the behavior of the perilymph and endolymph in the scala media, the scala vestibuli and the scala tympani. As the fluids are incompressible is used the conservation of momentum to describe their movement and the divergence theorem for to get the boundary conditions for small amplitudes ignoring the nonlinear terms. The mechanical behavior of the basilar membrane is analyzed as a system of forced damped harmonic oscillators concatenated from the model proposed by Lesser and Berkeley, considering that the wave equation that described the motion on the basilar membrane is the boundary condition of the difference pressure on each point in the membrane and that the physical characteristics of mass, stiffness and damping along of the membrane have different values that depended of the distance from the apex to the helicotrema.

The new model of resonance analysis considers that the system is excited by a periodic and complex external force caused by the vibrations transmitted within the cochlea through the oval window, this force is the solution for the equation of damped forced oscillator. Therefore is considered that the displacement and amplitude are also complex, next is defined the complex impedance mechanical of input to the system as the sum of the real part given by damping and an imaginary part provided by the mechanical reactance, expressed in polar form the terms of the components of magnitude and phase. The Euler's identity is used for separate the real part of the imaginary and to obtain an expression for the amplitude of the systems that depends only of the physical characteristics of mass, stiffness and damping along the basilar membrane and the frequency of exciting force.

This new model had the advantage over the existing models of determined the distance along the basilar membrane in which is present the maximum amplitude for all the interval of frequencies in the human hearing. For their evaluation is realized the comparison with the models of the cochlea developed by Peterson and Bogert (numerical integration method), Allen (Analysis by the Green's function), Neely (finite difference method) and the experimental measurements of Bekesy, in each experiment are used the physical parameters of the basilar membrane and the frequencies of evaluation in each model. Finally is shows an application to speech recognition developing a new form of parametrization and its implementation in HTK, is proposed a filter array using the finite difference method and the analysis of resonance for determined the interval of frequencies for each filter. In all experiments satisfactory results are obtained.

Dedicatoria

...Para mi amada hija María Jazmín Jiménez Arzeta que es todo para mi en la vida.

Bienaventurados los puros de corazón, porque veran a Dios... Mateo V, 8.

...Para mi esposa Rosalva Arzeta Salas que me acompaña siempre con amor y comprensión en nuestro camino.

Entonces el Señor Dios hizo caer sobre el hombre un profundo sueño, y cuando este se durmió, tomó una de sus costillas y cerró con carne el lugar vacío. Luego, con la costilla que había sacado del hombre, el Señor Dios formó una mujer y se la presentó al hombre. El hombre exclamó: ¡Esta sí que es hueso de mis huesos y carne de mi carne! ... Genesis II, 21-23.

...Para mi mamá Zenaida Hernández Hidalgo y para mi papá José Mario Jiménez Rojas que siempre estan presentes para darme su apoyo y consejos en la vida.

Honra a tu padre y a tu madre, para que tengas una larga vida en la tierra que el Señor, tu Dios, te da... Exodo XX, 12.

...Para Dios por haberme dado la oportunidad de estar aqui y permitirme superarme intelectual y espiritualmente.

Entonces el Señor Dios modeló al hombre con arcilla del suelo y sopló en su nariz un aliento de vida. Así el hombre se convirtió en un ser viviente... Genesis II, 7.

Todo hombre debe decidir una vez en su vida, si se lanza a triunfar arriesgándolo todo o si se queda a contemplar el paso de los triunfadores... Anónimo.

Mario Jiménez Hernández 2013

Agradecimientos

En primer lugar quiero agradecer a mi Alma Mater el Instituto Politécnico Nacional el haberme brindado mi formación académica de enseñanza superior hasta esta estancia final de estudio doctoral.

De forma muy especial al Centro de Investigación en Computación por albergarme durante esta parte de mi vida. Quiero resaltar mi agradecimiento al Laboratorio de Procesamiento Digital de Señales por las facilidades brindadas para la realización de este trabajo de Tesis tanto en el aspecto de herramientas de sofware como de hardware y por el compañerismo recibido por parte de sus integrantes.

A continuación quiero dar un agradecimiento muy especial al Dr. Ricardo Barrón Fernández por los conocimientos que me proporcionó dentro y fuera del salón de clases y por sus comentarios brindados y sugerencias durante la elaboración de este trabajo de Tesis.

Deseo resaltar un profundo agradecimiento a mi asesor el Dr. Sergio Suárez Guerra por todo el apoyo brindado en el aspecto académico y administrativo que me proporcionó en mi estancia en el centro de investigación durante su jefatura del Laboratorio de Procesamiento Digital de Señales y en su actual coordinación del Programa de Doctorado en Ciencias de la Computación

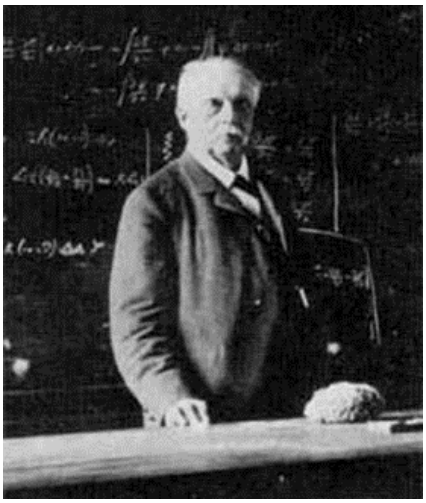
De forma muy significativa agradecer a mi asesor el Dr. José Luis Oropeza Rodríguez por sus conocimientos que me brindó durante mi trayectoria académica para la culminación de este trabajo, su guia, sugerencias y consejos durante este arduo camino.

Quiero dar un agradecimiento especial a mi centro de trabajo la Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional por las facilidades proporcionadas para poder continuar mis estudios de Doctorado.

Agradecer al Consejo Nacional de Ciencia y Tecnología por haberme dado el soporte económico para poder solventar mi etapa de formación doctoral.

Un agradecimiento de forma muy especial al Instituto Mexicano de Acústica y a la Acoustical Society of America (Sociedad Americana de Acústica) por su apoyo que me brindaron en los congresos celebrados durante mi estancia académica en el Centro de Investigación en Computación.

Por último agradecer amablemente al comité editorial y al personal de la Sociedad Mexicana de Ingeniería Biomédica por sus amables atenciones brindadas a mi persona.



...Von der Zerlegung der Klängle durch das Ohr

Est ist in den vorausgehenden Abschnitten schon mehrfach erwähnt worden, dass musikalische Klänge auch durch das menschliche Ohr allein, ohne dass irgend welche Unterstützung durch besondere Apparate nöthig wäre, in eine Reihe von Partialtönen zerlegt werden, die den einfachen pendelartigen Schwingungen der Luftmasse entsprechen, also in dieselben Bestandtheile, in welche die Bewegung der Luft auch durch mittönende elastische Körper zerlegt wird. Wir gehen jetzt daran, die Richtigkeit dieser Behauptung zu erweisen...

Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik

von H. Helmholtz 1877

Nomenclatura

Amplitud compleja	$\mathbf{A}$
Area	A
Cepstrum	$c(\tau)$
ídem. dominio Z	$C(z)$
Cepstrum complejo	$\widehat{c}_n$
ídem. dominio Z	$\widehat{C}(z)$
Coeficientes cepstrales	c_n
Coeficientes CLPC	α_m
Coeficientes de predicción lineal	a_k
Coeficientes MFCC	$c(n)$
Coeficientes PLP	$\theta(\Omega_i)$
Constante de elasticidad	k
Densidad	ρ
Distancia	x
Distancias de intervalos	$d(n)$
Distancia máxima	$dmax$
Distancia mínima	$dmin$
Desplazamiento de una onda	η
Escala de Mel	$\beta(f)$
Escala de Bark	$Bark(f)$
Espectro de potencia	$P(\omega)$
Espectro distorsinado	$\Omega(\omega)$
Espectro en potencia	$\Psi(\Omega)$
Error de predicción	e_n
ídem. cuadrático medio	E_n
Excitación discreta	$u(n)$
ídem. dominio Z	$U(z)$
Fase de la impedancia mecánica	Θ
Flujo irrotacional	ϕ
Frecuencia	f
Frecuencia máxima	$fmax$
Frecuencia mínima	$fmin$
Frecuencia de muestreo	f_m
Frecuencia de resonancia	f_R
Frecuencia angular	ω
Frecuencia angular de resonancia	ω_R
Fuerza	F
Fuente cuasiperiódica	$g(t)$
ídem. dominio de la frecuencia	$G(\omega)$
Filtro del tracto vocal en dominio Z	$H(z)$
Ganancia del filtro del tracto vocal	G
Impedancia mecánica compleja	$\mathbf{Z}$
Longitud	l

Masa	m
Magnitud de la impedancia mecánica	Z_m
Número de intervalos	$nint$
Presión	P
Denominador del modelo del tracto vocal en dominio Z	$A(z)$
Reactancia mecánica	X_m
Resistencia mecánica	R_m
Respuesta al impulso del tracto vocal	$h(t)$
ídem. dominio de la frecuencia	$H(\omega)$
Secuencia discreta	$x(n)$
ídem. dominio Z	$X(z)$
Segmento corto de voz	$s(t)$
ídem. dominio de la frecuencia	$S(\omega)$
Señal de voz	$x(t)$
ídem. dominio de la frecuencia	$X(\omega)$
ídem. dominio de la distancia	$X(d)$
Señal de voz muestreada	$s(n)$
ídem. predecida	$\widehat{s}$
ídem. dominio Z	$S(z)$
Superficie	S
Tiempo	t
Unidad de volumen	$\mathbf{u}$
Velocidad del sonido	c
Velocidad en un fluido	u
Velocidad de un volumen	U
Volumen	V

Glosario

Aparato fonador. Sistema generador de la voz formado por la cavidad bucal, la cavidad nasal, el tracto vocal, la laringe, la traquea y la glotis.

Ápice. Parte de la cóclea cercana a la ventana oval y la ventana redonda.

Bronquios. Conductos tubulares que conectan a la traquea con los pulmones.

Canal auditivo. Cavidad en el oído externo que une a la oreja con el tímpano.

Canales semicirculares. Sistema de canales en el oído interno que constituyen el sentido del equilibrio.

Cavidad bucal. Abertura corporal por la que se ingieren los alimentos.

Cavidad nasal. Cavidades que se abren en la cara a través de las dos aperturas nasales y se comunican con la faringe.

Cilios. Células en el órgano de Corti que transforman la energía hidráulica en biopoteciales eléctricos.

Cóclea. Elemento del oído interno en forma de caracol donde se realiza la selectividad en frecuencia de las señales acústicas.

Consonante. Fonema de una lengua oral originado por el cierre o estrechamiento del tracto vocal y por el acercamiento o contacto de los órganos de articulación.

Cuerdas Vocales. Labios membranosos ubicados dentro de la laringe cuya vibración da la característica del pitch a los sonidos.

Curvas de igual sonoridad. Respuesta del sistema auditivo entre la intensidad de una onda acústica y su frecuencia para diferentes umbrales de audibilidad.

Decibel. Unidad relativa logarítmica empleada en acústica para expresar la relación entre la magnitud que se estudia y una magnitud de referencia.

Diafragma. Músculo en forma de bóveda convexa ubicado entre la cavidad torácica y la cavidad abdominal.

Endolinfa. Líquido con alto contenido de potasio y bajo contenido de sodio ubicado en la Escala Media o Conducto Coclear.

Escala de Bark. Escala psicoacústica correspondiente a las primeras 24 bandas críticas del oído.

Escala de Mel. Escala perceptual de tonos musicales en intervalos equiespaciados en forma subjetiva por evaluación de observadores.

Escala media. Estructura tubular en forma de espiral llena de endolinfa ubicada entre las escalas vestibular y timpánica.

Escala timpánica. Estructura tubular en forma de espiral llena de perilinfa ubicada en la parte inferior de la cóclea debajo de las escalas vestibular y media.

Escala vestibular. Estructura tubular en forma de espiral llena de perilinfa ubicada en la parte superior de la cóclea arriba de las escalas media y timpánica

Estribo. Hueso en forma de estribo que conecta el oído medio con la ventana oval.

Faringe. Estructura en forma de tubo situada en el cuello que conecta a la nariz y la boca con la laringe.

Fonema. Unidad básica de sonido en una lengua oral.

Fonética. Estudio de los sonidos físicos del discurso humano.

Formante. Pico de intensidad en el espectro de un sonido.

Glotis. Espacio de la laringe donde se encuentran las cuerdas vocales.

Helicotrema. Orificio en la parte final de la cóclea que une a las escalas vestibular y timpánica.

Histograma. Representación gráfica de las frecuencias de una variable en forma de barras.

Imagen acústica. Imagen mental subjetiva que cada persona tiene ante un estímulo sonoro.

Intensidad acústica. Rapidez promedio del flujo de energía a través de un área unitaria y su dirección de propagación.

Laringe. Organo cilíndrico móvil ubicado en la parte superior de la traquea.

Lengua oral. Sistema de comunicación por voz con estructura sintáctica.

Lingüística. Estudio científico de la estructura de las lenguas orales.

Martillo. Hueso en forma de martillo que forma parte del oído medio.

Membragrama. Respresentación frecuencia distancia del comportamiento de la membrana basilar.

Membrana Basilar. Membrana situada en el interior de la cóclea donde se realiza por resonancia la selectividad en frecuencia de los sonidos.

Membrana de Reissner's. Memebrana que separa la escala vestibular de la escala media.

Nervio auditivo. Nervio craneal que comunica al cerebro con los sentidos del equilibrio y de la audición.

Oído externo. Parte exterior del sistema auditivo compuesta por la oreja y el conducto auditivo.

Oído interno. Cavidad en el hueso temporal del cráneo que incluye al vestíbulo y la cóclea.

Oído medio. Sistema de acoplamiento de impedancias acústicas entre el oído externo y el oído interno compuesto por el tímpano, el martillo, el yunque, el estribo y la trompa de eustaquio.

Organo de Corti. Membrana en el interior de la escala media que conecta a los cilios con el nervio auditivo.

Perilinfa. Líquido con características similares al suero que llena las escalas vestibular y timpánica.

Pitch. Frecuencia fundamental de los fonemas con movimiento glotal.

Potencia acústica. Cantidad de energía por unidad de tiempo emitida por una fuente de ondas acústicas.

Prosodia. Rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral.

Psicofísica. Rama de la psicología que estudia la relación entre la magnitud de un estímulo físico y la intensidad con la que éste es percibido por parte de un observador.

Pulmón. Órganos del aparato respiratorio donde se inhala y exhala el aire para realizar la hematosis.

Ruido. Sonido no deseado y molesto para un escucha.

Signo lingüístico. Combinación de un concepto y de una imagen acústica.

Sintaxis. Forma en que se combinan las palabras.

Semántica. Estudio del significado, sentido e interpretación de los signos lingüísticos.

Sistema auditivo. Conjunto de órganos que hacen posible la percepción de las ondas acústicas.

Tímpano. Membrana elástica que comunica las ondas acústicas que llegan

al canal auditivo con el oído medio.

Tracto vocal. Articuladores del aparato fonador que producen los armónicos de los sonidos del habla, siendo los principales: la lengua, el paladar y los labios.

Traquea. Conducto cartilaginoso que comunica a la laringe y los bronquios.

Trompa de Eustaquio. Estructura en forma de tubo que se extiende desde la caja del tímpano hasta la región nasofaríngea.

Umbral de audibilidad. Mínima presión necesaria para percibir un sonido senoidal en diferentes condiciones ambientales de intesidad acústica.

Ventana Oval. Membrana ubicada en el inicio de la cóclea que transmite las ondas acústicas provenientes del estribo a la escala vestibular.

Ventana Redonda. Membrana ubicada al inicio de la cóclea en donde se igualan las presiones generadas entre la escala timpánica y la trompa de Eustaquio.

Vestíbulo. Órgano del oído interno formado por los canales semicirculares y la escala vestibular.

Vocal. Fonema de una lengua oral que se pronuncia con el tracto vocal abierto y con el movimiento de la glotis.

Voz. Sonido generado por el aparato fonador.

Yunque. Hueso con forma de yunque que une al tímpano con el oído medio.

Índice general

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

El objetivo de la ciencia es una comprensión tan completa como sea posible de la conexión entre las experiencias sensoriales en su totalidad y el logro de ese objetivo mediante el uso de un mínimo de conceptos primarios y de relaciones...
Albert Einstein. Física y realidad, Princeton 1936.

El presente capítulo describe primero los antecedentes del modelado mecánico acústico del oído interno mostrando las diferentes soluciones existentes y los antecedentes del modelado del sistema auditivo aplicados a los procesos de reconocimiento de voz. Posteriormente se describe el planteamiento de solución de este trabajo de Tesis al problema del modelado del comportamiento mecánico acústico de la cóclea usando análisis por resonancia y su aplicación al problema de la parametrización de la señal de voz para procesos de reconocimiento. Por último se presentan la justificación del trabajo, la hipótesis a partir de la cual se plantea su solución, el objetivo general, los objetivos especificos, los alcances del trabajo y el contenido restante de la tesis.

1.1. Modelado mecánico acústico del oído interno

La primera teoría mecánica de la cóclea basada en la hidrodinámica fue propuesta por Peterson y Bogert en 1950, consideró a la cóclea como un sistema de dos canales que varían en forma similar en su sección transversal y separados por una membrana elástica con constantes dinámicas variables [Pet50]. Para su modelado se utilizaron los parámetros reportados en los trabajos experimentales de Békésy en 1934 [Bek60]. En los años siguientes se desarrollaron muchas teorías acerca de la mecánica de la cóclea, pero en 1971 Rhode realizó mediciones físicas apoyadas en la fisiología del sistema auditivo y las teorías que fueron propuestas anteriormente resultaron ser inadecuadas [Rho71] [Rho74].

Posteriormente en 1972 Lesser y Berkeley formularon un modelo que se ajustó a todas las observaciones reportadas, modelando a la cóclea como un sistema de mecánica de fluidos y a la membrana basilar como un sistema de osciladores armónicos forzados concatenados [Les72].

En 1976 Allen utiliza el modelo de Lesser y Berkeley para obtener los parámetros de la membrana basilar utilizando la función de Green, logrando obtener un conjunto de parámetros más aproximados de su comportamiento [All77]. Un trabajo posterior fue desarrollado en 1981 por Neely, en el cual se propone un modelo matemático en dos dimensiones de la cóclea, y su solución numérica empleando aproximaciones por diferencias finitas usando la ecuación de Laplace, obteniendo hasta estos momentos los mejores parámetros de la respuesta mecánica de la cóclea [Nee81].

La solución al modelo de la membrana basilar como un sistema de osciladores armónicos forzados, ha sido propuesta en forma numérica a partir del modelado de flujo de potencial por series de Fourier por Lesser y Berkeley en 1972 [Les72]. Posteriormente en 1974 Siebert generaliza la solución de Lesser y Berkeley considerando una fuerza mecánica en los dos extremos de la membrana basilar [Sie74], una solución similar fue encontrada en 1981 por Peskin [Pes81]. Sin embargo los estudios posteriores consideraron la forma de la membrana basilar para dar solución al modelo, destacando los estudios en 1984 de Rhode [Rho84], en 1985 de Hudspeth [Hud85] y en 1996 de Boer [Boe96]. En años recientes se ha dado solución al modelo considerando modelos de espacio estado, en 2007 por Elliott et al. [Ell07] y en 2008 por Ku et al. [Ku08].

1.2. Modelado del sistema auditivo en reconocimiento de voz

La respuesta del sistema auditivo ha sido utilizada para la parametrización óptima de la señal de voz para su evaluación en procesos de reconocimiento [Rab93] [Ben08]. Sin embargo los primeros modelos para el reconocimiento de voa fueron desarrollados a partir de la caracterización del proceso de generación de la voz [Cas89] [Ber00], siendo dos los principales: el Cepstrum [Nol64] y el análisis predictivo lineal o LPC [Ata71]. A pesar de que el estudio de los modelos de generación de la señal de voz permitieron obtener vectores de coeficientes representativos que proporcionaron resultados satisfactorios al ser utilizados en las metodologías de reconocimiento, fueron los parámetros obtenidos a partir de la respuesta del sistema auditivo a la percepción de la voz los que mejoraron notablemente los resultados en la parte del reconocimiento [Wai90] [Gol11].

Una forma eficiente para extraer un vector de coeficientes fue desarrollada a partir de criterios perceptuales del sistema auditivo, los cuales demuestran que la percepción de los tonos en los humanos no está dada en una escala lineal y por lo tanto se propone utilizar la escala de Mel [Yos06]. Se obtiene una representación definida del Cepstrum de una señal ventaneada en el tiempo a partir de su transformada de Fourier, distibuyendo el comportamiento espectral

en bandas de frecuencia mediante un banco de filtros con los que se calcula el promedio del espectro alrededor de cada frecuencia central. A estos parámetros se les denomina MFCC y tienen como ventaja poder establecer un criterio de limitación en banda comunmente útil para rechazar frecuencias no deseadas o evitar la construcción de filtros en regiones de frecuencia en los cuales no existe energía útil en la señal [Dav80].

Un método alterno fue propuesto usando la implementación de la predicción lineal perceptual, tomando como base tres conceptos de la psicofísica de la audición para derivar una estimación del espectro presente en la audición: la resolucion espectral de banda crítica, las curvas de igual sonoridad y la potencia en la intensidad de la señal. El espectro en el sistema auditivo es entonces aproximado por un modelo autoregresivo todo polos obteniendo los parámetros PLP, siendo este tipo de análisis más consistente con la audición humana. En este planteamiento se integra la teoría de la percepción con base en la escala de Bark proporcionando resultados más satisfactorios que con todas las metodologías ya existentes [Her90].

1.3. Planteamiento del problema

El sistema auditivo es el organo sensorial humano más importante después de la visión. Está dividido en tres elementos principales: el oído externo, el oído medio y el oído interno [Kin00]. Fisiológicamente existen casos clínicos de daño parcial o total en el oído externo y alteraciones patológicas en el oído medio donde los sujetos de estudio continuan percibiendo señales debido a las vibraciones trasmitidas a través del sistema óseo al oído interno, siendo la cóclea el elemento que conecta a este órgano con el cerebro a partir del nervio auditivo [Yos06]. El desarrollo de los implantes cocleares [Fur92] demuestra que la cóclea es el elemento principal de la audición, de ahí que el modelado de este órgano sea esencial para la comprensión de la audición.

Los modelos existentes de la cóclea han demostrado ser sólo eficientes para ciertos intervalos de frecuencia, teniendo algunos modelos problemas en las altas frecuencias cercanas al apíce y otros en las bajas frecuencias cercanas al helicotrema, lo cual es debido a la técnicas de modelado utilizadas en su desarrollo [Kee08]. Los modelos más aproximados del comportamiento físico de la cóclea han dado soluciones que dependen de la forma de la envolvente de la onda propagada en su interior [Les72], sin considerar la relación de dependencia puntual entre la frecuencia y la distancia a lo largo de la membrana basilar la cual está determinada por sus características físicas de masa, resistencia mecánica y constante de elasticidad para todo el intervalo de frecuencias del sistema auditivo humano [Bek60]. En este trabajo de Tesis se propone una solución a las deficiencias antes mencionadas usando el análisis por resonancia del modelo de la membrana basilar como un sistema de osciladores armónicos forzados propuesto por Lesser y Berkeley [Jim10a] [Jim10b].

Las soluciones desarrolladas hasta la actualidad al problema del reconocimiento de voz se encuentran distantes de las espectativas cientifícas de la interfaz

hombre máquina propuesta por las tendencias de la inteligencia artificial [Eco93] [Del00] [Qua02]. Los modelos de parametrización que mejores resultados proporcionan son los basados en la percepcion auditiva, sin embargo el modelado del oído interno ha sido propuesto sólo para el análisis de señales de voz a partir de arreglos de filtros que emulan el funcionamiento de la cóclea y no han sido probados en etapas de reconocimiento [Fur92]. En este trabajo de tesis se da una solución al problema de la parametrización de la señal de voz usando el modelo de análisis por resonancia de la cóclea para crear un banco de filtros que modelen su respuesta en forma similar a los MFCC [Jim13].

1.4. Propuesta de solución y nuevas aportaciones científicas

En este trabajo de Tesis se propone el empleo de una solución alternativa del modelo de la mecánica de fluidos de la cóclea propuesto por Lesser y Berkeley [Les72] y su solución del comportamiento de la membrana basilar como un sistema de osciladores armónicos forzados concatenados utilizando el análisis por resonancia, considerando únicamente los parámetros de masa, resistencia mecánica y constante de elasticidad a lo largo de la membrana basilar [Jim12]. La solución desarrollada proporciona la relación frecuencia distancia del sistema con la cual se determina la distancia a lo largo de la membrana basilar donde se presenta la mayor amplitud para una frecuencia de excitación específica. Siendo esta solución concordante en su totalidad con la teoría de los puntos de audición de Békésy.

El planteamiento de la solución del modelo de la membrana basilar de Lesser y Berkeley usando análisis por resonancia considera que el sistema del oscilador armónico forzado amortiguado es excitado por una fuerza externa periódica y de forma compleja, la cual es producida por las vibraciones transmitidas a través del estribo al interior de la cóclea por la ventana oval, siendo esta fuerza la solución de la ecuación del oscilador armónico forzado amortiguado y por lo tanto el desplazamiento y la amplitud también se consideran complejos. Posteriormente se define la impedancia mecánica compleja de entrada del sistema como la suma de la parte real dada por la resistencia mecánica y la parte imaginaria dada por la reactancia mecánica pudiendo expresar en forma polar los términos de las componentes de magnitud y fase. A continuación se utiliza la identidad de Euler para separar la parte real de la ecuación de la parte imaginaria y se obtiene una expresión para la amplitud del sistema que depende sólo de las características físicas de masa, constante de elasticidad y resistencia mecánica a lo largo de la membrana basilar y de la frecuencia de la fuerza excitadora del sistema.

Para la evaluación del modelo de análisis por resonancia se hace la comparación con los resultados obtenidos en los trabajos de Peterson y Bogert [Pet50], Allen [All77], Neely [Nee81] y los obtenidos en forma experimental por Békésy [Bek60], utilizando en todos los experimentos las mismas frecuencias y los mis-

mos parámetros de la membrana basilar reportados en cada trabajo con el objetivo de hacer la comparación entre los datos obtenidos del análisis por resonancia y los datos reportados por cada investigador. Su aportación respecto a las diferentes soluciones ya encontradas, es obtener un valor unívoco entre la distancia a lo largo de la membrana basilar y cada frecuencia particular de excitación al oído interno en todo el intervalo de la audición humana, el cual sólo depende de las características físicas a lo largo de la membrana basilar.

Usando la solución desarrollada se propone una metodología de parametrización de la señal de voz partiendo de la elección de un intervalo de trabajo útil entre dos frecuencias límites. A partir de la relación frecuencia distancia del análisis por resonancia se obtienen las distancias límites correspondientes sobre la membrana basilar para las frecuencias inferior y superior, a continuación se divide esta distancia en un número de intervalos equidistantes usando el método de diferencias finitas. Posteriormente las distancias de los intervalos son transformadas al dominio de la frecuencia para construir un arreglo de filtros triangulares similares a los empleados por los coeficientes MFCC, finalizando el proceso de parametrización y reconocimiento en forma similar a esta metodología [Jim13]. El arreglo de filtros MFCC modela la respuesta del sistema auditivo con base en la escala de Mel y el arreglo de filtros propuesto modela el comportamiento mecánico acústico de la cóclea, para su evaluación en procesos de reconocimiento de voz se implementa la solución propuesta en la herramienta HTK [You09].

1.5. Justificación

La investigación del comportamiento mecánico acústico del oído interno es parte fundamental para el entendimiento de la forma en que la cóclea proporciona la información al cerebro a partir de las señales acústicas, debido a esto la obtención de un modelo más aproximado de su comportamiento frecuencia distancia permite realizar avances en este campo de la ciencia, complementando los estudios previos y sirviendo como base para la resolución de diferentes problemas tales como la parametrización de la señal de voz para su aplicación en procesos de reconocimiento y la determinación de las distancias de los electrodos en un implante coclear.

1.6. Hipótesis

Debido a que la cóclea es el elemento del oído interno en el cual se realiza la transducción de energía mecánica hidráulica en biopotenciales eléctricos, es posible obtener a partir de los parámetros físicos de la membrana basilar un modelo que emule su comportamiento en forma unívoca pudiendo obtener una relación frecuencia distancia usando análisis por resonancia.

1.7. Objetivo general

Obtener un modelo mecánico acústico de la respuesta de la cóclea que dependa solo de las características físicas de la membrana basilar usando análisis por resonancia.

1.8. Objetivos específicos

- Realizar un estudio de los modelos mecánico acústicos del oído interno basados en la mecánica de fluidos y que modelan a la membrana basilar como un oscilador armónico forzado amotiguado.
- Desarrollar un modelo mécanico acústico del oído interno usando análisis por resonancia que represente el comportamiento físico de la membrana basilar considerando sus características físicas.
- Comprobar los resultados del modelo desarrollado con modelos existentes en la literatura.
- A partir del modelo propuesto desarrollar una forma de parametrización de la señal de voz mediante un arreglo de filtros que represente el comportamiento de la cóclea.
- Aplicar la parametrización desarrollada en procesos de reconocimiento de voz comparando los resultados obtenidos con las metodologías existentes en la literatura.

1.9. Alcances del trabajo

Debido a que el modelo desarrollado emula el comportamiento mecánico acústico de la cóclea a partir de las características físicas de la membrana basilar, este trabajo de Tesis sólo considera este tipo de modelos y no contempla los modelos fenomenológicos o de procesamiento de señales biológicas de la respuesta del oído interno. De los modelos fenomenológicos sólo se presenta en el estado del arte el modelo de la respuesta del oído interno desarrollado por Ghitza [Fur92] ya que ha sido el único utilizado en el análisis de señales de voz.

De todos los modelos mécanico acústicos desarrollados por diversos investigadores sólo se presentan en el estado del arte los más significativos y mejor fundamentados, evitando mostrar modelos que únicamente fueron propuestas o extensiones de modelos que no brindaron resultados satisfactorios, acotando el trabajo a los modelos que tienen como fundamento físico la mecánica de fluidos y presentan soluciones particulares basadas en este tipo de modelado.

Los experimentos desarrollados se basan en un corpus de voz para el idioma español con hablantes masculinos, las grabaciones se realizaron dentro de un laboratorio de informática, siendo el ruido presente durante las pruebas desarrolladas no significativo para la inteligibilidad de los comandos pronunciados.

1.10. Contenido restante de la Tesis

En el capítulo dos se presenta la fisiología básica del sistema auditivo, posteriormente se describen los principales modelos mecánico acústicos del oído interno mostrando sus características físicas, a continuación se muestra en forma cronológica el estado del árte del análisis y del reconocimiento de voz haciendo énfasis en las metodologías de parametrización, después se describe el único modelo del oído interno que ha sido utilizado en procesos de análisis de la señal de voz, por último se plantea la propuesta de solución de este trabajo de Tesis del modelado del oído interno y su aplicación al problema de la parametrización de la voz para su evaluación en procesos de reconocimiento.

En el capítulo tres se muestra el desarrollo de la propuesta de solución de este trabajo de Tesis, primero se detalla la mecánica de fluidos en la cóclea propuesta por Lesser y Berkeley mostrando las condiciones de frontera del modelo [Les72], después se presenta la solución del comportamiento de la membrana basilar como un sistema de osciladores armónicos forzados concatenados, a continuación se detalla la propuesta de solución al modelo antes descrito utilizando análisis por resonancia y por último se describe la propuesta de parametrización de la señal de voz mediante un arreglo de filtros desarrollado a partir del método de diferencias finitas y la solución propuesta.

En el capítulo cuatro se describen los experimentos para la evaluación del análisis por resonancia, primero se hace la comparación con los modelos de Peterson y Bogert [Pet50], Allen [All77], Neely [Nee81] y los resultados experimentales de Békésy [Bek60]. A continuación se muestran los experimentos de análisis de señales de voz usando la representación frecuencia distancia del comportamiento de la membrana basilar *membragrama* usando los parámetros de Neely. Posteriormente se presentan los resultados del arreglo de filtros desarrollado y su implementación en procesos de reconocimiento de voz en HTK [You09].

En el capítulo cinco se presentan las aportaciones científicas del trabajo desarrollado, las conclusiones obtenidas y los trabajos futuros. A continuación se muestra la bibliografía consultada en forma detallada y por último se presentan en los apéndices aspectos matemáticos del modelado del oído interno y del reconocimiento de voz, junto con los resúmenes de las ponencias en congresos y los artículos publicados.

Capítulo 2

Estado del arte

Si he logrado ver más lejos, ha sido porque he subido a hombros de gigantes...
Isaac Newton. Carta a Robert Hooke, Londres 1676.

El presente capítulo muestra la fisiología del sistema auditivo y los principales modelos mecánico acústicos del oído interno. A continuación se describen los fundamentos del análisis, parametrización y reconocimiento de voz. Posteriormente se muestra el modelado del oído interno usado en el análisis de la señal de voz. Por último se plantea la propuesta de este trabajo de Tesis al problema del modelado del oído interno y la parametrización de la señal de voz para su reconocimiento.

2.1. Fisiología del sistema auditivo

El sistema auditivo es el encargado de procesar las señales acústicas para su reconocimiento en el cerebro, realiza una transformación de la señal de entrada del dominio del tiempo al dominio de la frecuencia para señales entre 20 Hz y 20 kHz, el conjunto de frecuencias y el nivel de presión sonora no son percibidos con la misma sensibilidad por los humanos, las altas y las bajas frecuencias se perciben con menor intensidad que las frecuencias medias, siendo la zona cercana a los 3 kHz la de mayor sensibilidad. Esta variación fué reportada por primera vez por Fletcher y Munson en un artículo para la JASA en 1933 titulado *Loudness, Its Definition, Measurement and Calculation* [Fle33], mostrando que la respuesta del sistema auditivo a estímulos externos presenta un comportamiento logarítmico. En la figura 2.1 se muestran las curvas de igual sonoridad para diferentes presiones acústicas del sistema auditivo y los umbrales de audibilidad [Kin00].

Fisiológicamente el sistema auditivo está dividido para su estudio en tres subsistemas, el oído externo, el oído medio y el oído interno. El oído externo está formado por la oreja y el canal auditivo, el cual se conecta en su extremo

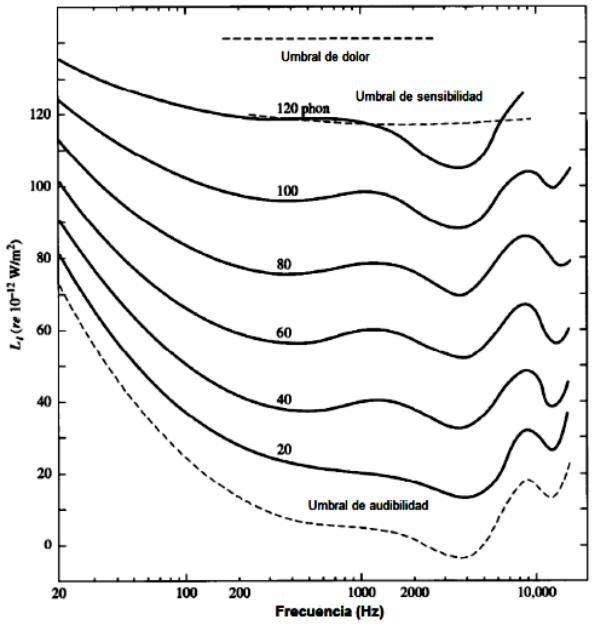


Figura 2.1: Umbrales de audibilidad. [Kin00]

final con el tímpano, siendo este elemento el que separa al oído externo del oído medio. El oído medio está compuesto por el tímpano y una cadena de tres huesos concatenados, llamados por su forma física martillo, yunque y estribo, estando el tímpano conectado al martillo y el esribo conectado en su parte final a la ventana oval ubicada en el principio de la cóclea. El oído interno está compuesto por el vestíbulo y la cóclea, en la figura 2.2 se muestran los elementos que forman al sistema auditivo [Kin00].

El oído externo percibe las ondas acústicas por medio de la oreja y las canaliza a través del conducto auditivo hasta el tímpano, éste funciona como un resonador al reforzar las ondas acústicas entre los 2.5 kHz y los 4.5 kHz, presentando una ganancia de presión acústica en el tímpano de dos a cuatro veces más respecto a la presión de entrada del sistema, lo cual mejora la sensibilidad en este intervalo de frecuencias. Sirve como acoplador térmico al mantener uniformes la temperatura y la humedad del aire en su vecindad, protegiendo la sensibilidad del tímpano y permitiendo la localización de fuentes de sonido con un alto grado de selectividad direccional debido a la estructura y forma física de la oreja.

El oído medio realiza un acoplamiento de impedancias entre la presión acústica proveniente de las ondas recibidas por el oído externo y las vibraciones mecánicas canalizadas hacia el oído interno, protegiendo al oído interno de intensidades de presión elevadas que llegan al tímpano. El tímpano es una membrana

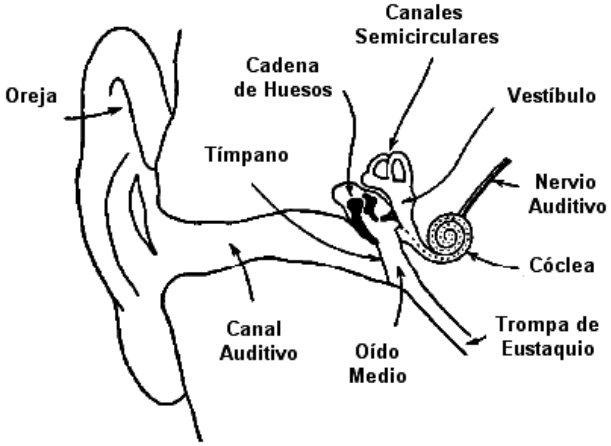


Figura 2.2: Elementos del sistema auditivo. [Kin00]

elástica relativamente rígida en forma de cono que vibra con los cambios de presión que llegan por el conducto auditivo, con las bajas frecuencias vibra casi en su totalidad y con las altas frecuencias únicamente en diferentes partes de su membrana, se encuentra acoplado con la cadena de huesos a través del martillo.

La cabeza del martillo se mueve sobre la superficie articulatoria del yunque, el cual en su parte inferior se enlaza con la cabeza del estribo, la base del estribo está unida al oido interno mediante la ventana oval. La cadena de huesos actúa como un conjunto de niveladores que incrementan la fuerza a expensas de la velocidad, transmitiendo sólo la energía requerida por la ventana oval, aumentan catorce veces la presión que llega a la ventana oval con relación a la presión que tenian las ondas que llegan al tímpano y cuando la fuerza es grande se produce una disminución de la intensidad. El oído medio está abierto hacia la faringe por medio de la trompa de Eustaquio la cual es una cavidad llena de aire. Este conducto sirve para igualar la presión del aire contenido en el oído medio con la presión del aire exterior, sin esta condición la membrana del tímpano no podría vibrar en perfectas condiciones [Qui99].

En el oído interno el vestíbulo está constituido por tres canales semicirculares y una cavidad vestibular, su función es la detección del movimiento y la aceleración en el sentido del equilibrio, en la audición no tiene participación, todos los procesos relacionados con la audición ocurren dentro de la cóclea. En la cóclea se convierten las vibraciones mecánicas provenientes del oído medio en biopotenciales eléctricos que son interpretados en el cerebro, la forma y estructura de la cóclea son la de un caracol constituido por un elemento tubular de aproximadamente 35 mm de longitud que disminuye su diámetro a partir de la ventana oval, enrollándose dos veces y media.

La cóclea está alojada en los huesos del cráneo y está dividida en forma lon-

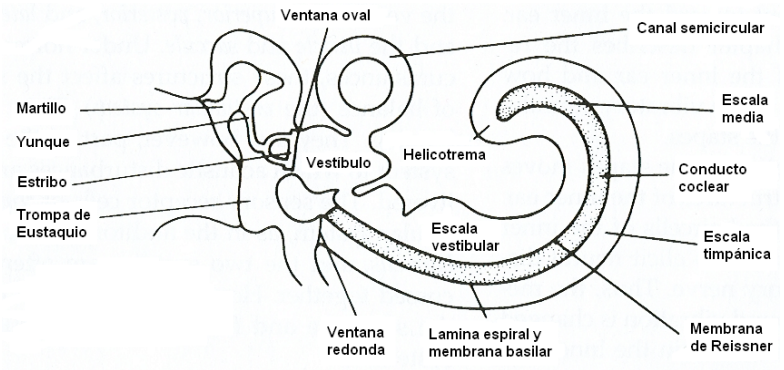


Figura 2.3: Elementos del oído interno. [Yos06]

gitudinal en tres compartimientos, la escala vestibular en la parte superior, la cual está comunicada libremente con el vestíbulo y está en contacto con la ventana oval, la escala timpánica que se comunica con la ventana redonda, ambas se comunican en la parte final de la cóclea por medio del helicotrema, en su interior están llenas de un líquido claro y viscoso llamado perilinfa, el tercer compartimento es la escala media, la cual está llena de un fluido llamado endolinfa que tiene una alta concentración de potasio y una baja concentración de sodio. Las tres escalas se enrollan en conjunto a lo largo de la espiral preservando su orientación espacial, en la figura 2.3 se muestran los elementos del oído interno y el esquema de la cóclea con las posiciones relativas de los compartimientos internos de la escala vestibular y la escala timpánica [Yos06].

En la parte central se encuentra el conducto coclear que está lleno de endolinfa, la membrana de Reissner's separa la escala vestibular del conducto coclear y de la escala media, la cual a su vez está separada desde la escala timpánica por la lámina espiral y la membrana basilar. La membrana basilar separa el conducto coclear de la escala timpánica, por un lado se une al ligamento espiral que envuelve la pared externa de la cóclea, la membrana basilar va ensanchándose a medida que se aproxima al helicotrema.

Las vibraciones de sonido que son transmitidas a través del oído medio son canalizadas a través de la ventana oval dentro de la escala timpánica, las ondas resultantes dentro de la perilinfa viajan a lo largo de la escala vestibular, creando ondas complementarias en la membrana basilar y la escala timpánica, el helicotrema ecualiza la presión local en los dos compartimientos. Debido a que la perilinfa es esencialmente un fluido incompresible, es necesario para la escala timpánica tener también una abertura análoga a la ventana oval, logrando con esto que la cantidad de movimiento de masa se conserve durante la propagación del movimiento ondulatorio, esta abertura es la ventana redonda, con esta característica se compensa el movimiento de la masa del fluido que se genera en la ventana oval con el de la ventana redonda, en la figura 2.4 se muestra una

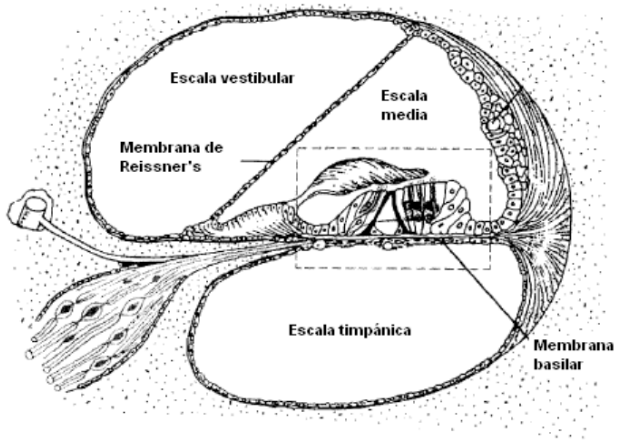


Figura 2.4: Seccción transversal de la cóclea. [Kee08]

vista de la sección transversal de la cóclea [Kee08].

La transducción de las ondas acústicas en impulsos eléctricos se realiza en el órgano de Corti, el cual está colocado sobre la parte superior de la membrana basilar. Las células ciliadas o cilios en el órgano de Corti están sujetas mediante filamentos a la membrana tectorial que se encuentra en la parte superior del órgano de Corti y el otro extremo en la membrana basilar, estando el sistema constituido por cerca de 25000 células ciliadas. Las ondas en la membrana basilar generan una fuerza sobre los cilios que causa un cambio en su potencial que es transmitido a través de los nervios auditivos directamente al cerebro para realizar la interpretación de las diferentes frecuencias, en la figura 2.5 se muestra un diagrama de la membrana basilar incluyendo al órgano de Corti y los cilios [Kee08].

Las vibraciones que se propagan sobre la membrana basilar fueron estudiadas por primera vez por Békésy [Bek60], sus estudios demostraron que la forma de onda tiene una amplitud envolvente que es una función de dos dimensiones entre la distancia desde la base al helicotrema y la frecuencia, cuando la frecuencia se incrementa la cresta de la envolvente se mueve hacia la base de la cóclea donde se encuentra la ventana oval y la ventana redonda, cuando la frecuencia se decrementa la cresta de la envolvente se mueve hacia el helicotrema. Por lo tanto la cóclea permite identificar las frecuencias que conforman una onda de sonido, la parte de la membrana basilar cercana a la base responde a las altas frecuencias cercanas a los 20 kHz y la parte junto al helicotrema responde a las bajas frecuencias cercanas a los 20 Hz. En la figura 2.6 se pueden observar las formas de onda sobre la membrana basilar y su envolvente, las líneas continuas muestran la deflexión de la membrana basilar en tiempos sucesivos 1, 2, 3, 4 y la línea punteada representa la posición de la envolvente de la onda sobre la membrana basilar, el pico de la envolvente es dependiente de la frecuencia del

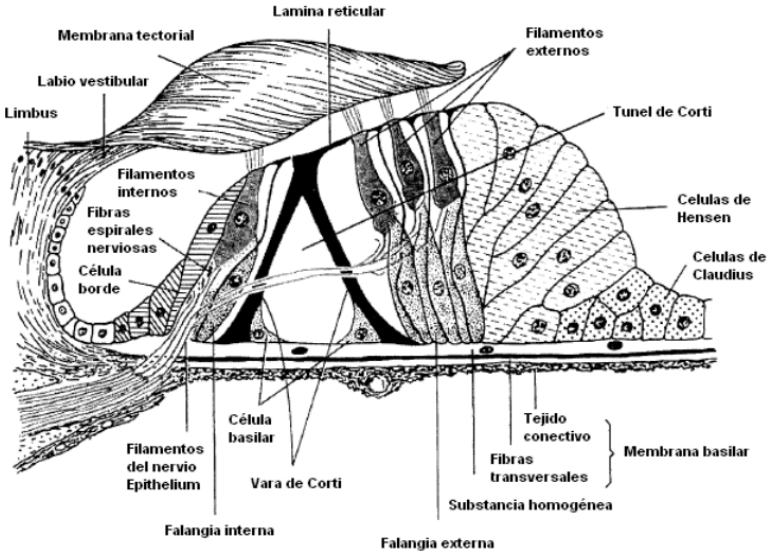


Figura 2.5: Elementos de la membrana basilar. [Kee08]

estimulo [Yos06].

Otra característica de la propagación de las ondas sobre la membrana basilar es que su velocidad decrementa conforme la onda se desplaza a lo largo de la membrana, lo cual da como resultado un decremento continuo en fase y un aparente incremento en la frecuencia, en la figura 2.7 se muestran las curvas de respuesta de amplitud y fase de onda sobre la membrana basilar para cuatro diferentes frecuencias [Yos06].

El ancho de la cóclea decrece desde su base hasta el helicotrema, sin embargo el ancho de la membrana basilar se incrementa en esta dirección, la resistencia que presenta la membrana basilar al movimiento es debida a su característica de elasticidad al momento de doblarse y su rigidez decrece en forma exponencial desde la base hasta el ápice. Debido a que cada parte de la membrana basilar responde al valor máximo de la cresta de la onda envolvente y que la frecuencia de la onda envolvente se incrementa en forma aparente en cada punto específico de la membrana, se considera que el mecanismo de la cóclea determina la frecuencia de la señal de entrada a partir del lugar donde la membrana basilar presenta una amplitud máxima lo cual es la teoría de Békésy de los puntos de audición, en la figura 2.8 se muestran sus resultados experimentales [Kee08].

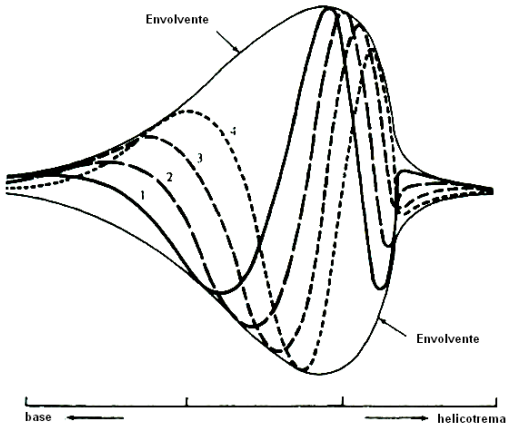


Figura 2.6: Envolvente de onda en la membrana basilar. [Yos06]

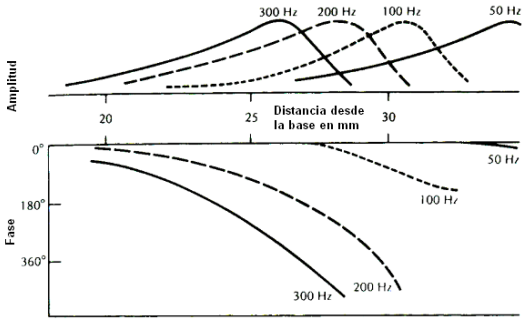


Figura 2.7: Amplitud y fase de onda en la membrana basilar. [Yos06]

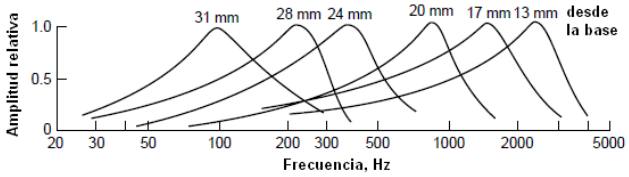


Figura 2.8: Relación frecuencia distancia en la membrana basilar. [Kee08]

2.2. Modelado físico del oído interno

El primer modelo físico del oido interno fue propuesto por Helmholtz en 1885 [Hel54], considera que la membrana basilar es una membrana resonante que vibra en lugares específicos y que dependiendo de la frecuencia de excitación externa se habilita sólo a un conjunto específico de cilios. El sistema es una solución particular de la ecuación de onda para un sistema físico de dos dimensiones. Este modelo presenta la desventaja de considerar las características físicas uniformes a lo largo de la membrana basilar, sin embargo al modelar su comportamiento como un sistema de resonancia proporciona una aproximación de su comportamiento.

A finales de los años 40 y comienzos de los años 50 se desarrollarón las primeras teorías del funcionamiento de la cóclea basadas en la hidrodinámica de fluidos, en 1950 Peterson y Bogert estudian la hidrodiñamica en el interior de la cóclea y publican sus resultados para la JASA en un artículo titulado *A Dynamical Theory of the Cochlea* [Pet50], ese mismo año Ranke presenta sus resutados de la operación de la cóclea en un artículo para la JASA titulado *Theory of Operation of the Cochlea: A Contribution to the Hidrodinamycs of the Cochlea* [Ran50], al mismo tiempo Zwislocki desarrolla una teoría acústica de la cóclea publicando sus resultados para la JASA en un artículo titulado *Theory of the Acoustical Action of the Cochlea* [Zwi50], posteriormente en 1953 Zwislocki hace una revisión de los modelos de la cóclea desarrollados hasta ese momento y lo publica para la JASA en un artículo titulado *Review of Recent Mathematical Theories of Cochlear Dynamics* [Zwi53].

De los trabajos antes descritos el modelo de Peterson y Bogert es considerado una de las mejores aproximaciones del comportamiento de la cóclea, su estudio modela a la cóclea como un sistema de dos canales que varían en forma similar en su sección transversal, estando separados por una membrana elástica con constantes dinámicas variables las cuales son dependientes de la distancia entre la ventana oval y el helicotrema. Para caracterizar su modelo utilizaron los datos experimentales reportados por Békésy, usando el mismo valor de la constante de elasticidad para establecer las condiciones límites de su modelo y determinar las velocidades y fases de las ondas a lo largo de la membrana basilar, siendo el valor de la constante de elasticidad:

$$k(x) = 1{,}72 \cdot 10^9 e^{-2x} \tag{2.1}$$

La masa de la membrana basilar está en función de la forma física del ducto coclear, presentando un valor constante de $m = 0{,}143$ en g/cm^2. En la figura 2.9 se muestra la gráfica de los resultados de Peterson y Bogert para las frecuencias de 31.6 Hz, 100 Hz, 316 Hz, 1000 Hz, 3160 Hz y 10000 Hz, siendo el eje x la distancia de excitación a lo largo de la membrana basilar desde la ventana oval hasta el helicotrema en cm y el eje y la respuesta en amplitud de la membrana basilar en dB [Pet52].

Este modelo presenta la ventaja de considerar que los dos canales están interconectados en sus extremos y que su estructura es una forma rígida excepto

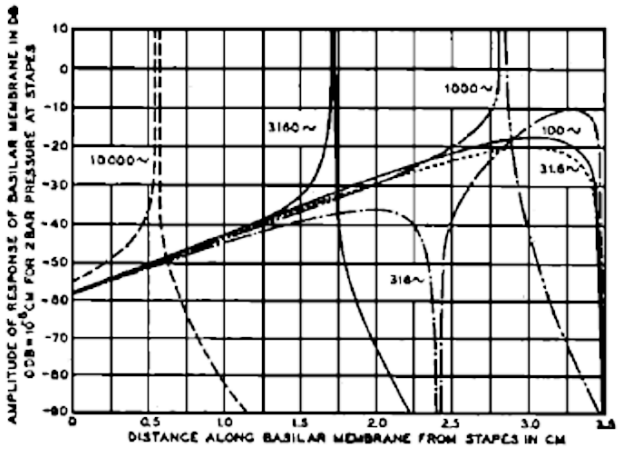


Figura 2.9: Respuesta de la membrana basilar por Peterson y Bogert. [Pet52]

en las áreas correspondientes a la ventana oval y la ventana redonda, sin embargo tiene la desventaja de que la ecuación del movimiento incluye el efecto de la membrana basilar, considerando que el sistema es no disipativo y que por lo tanto no presenta resistencia mecánica.

En 1960 Békésy hace un compendio de todas sus observaciones del comportamiento coclear y las publica en un libro titulado *Experiments in Hearing* [Bek60]. A comienzos de la década de los años setentas Lesser y Berkeley formulan un modelo de la mecánica de fluidos de la cóclea que se ajusta a todas las observaciones reportadas, publicando sus resultados para el *Journal of Fluid Mechanics* en un artículo titulado *Fluid Mechanics of the Cochlea. Part I* [Les72]. Para la determinación de las características físicas de la membrana basilar tomaron como base los trabajos de Peterson y Bogert [Pet50] y Zwislocki [Zwi53], su modelo considera por simplicidad únicamente las escalas vestibular y timpánica. La solución propuesta en su artículo del comportamiento de la membrana basilar está dada mediante una solución numérica empleando series de Fourier la cual es descrita en el apéndice A. En este modelo se considera que la constante de elasticidad a lo largo de la membrana basilar varía en $dyna \cdot seg/cm^2$ en dos posibles formas:

$$k(x) = 10^7 e^{-1,5x} \tag{2.2}$$

$$k(x) = 10^9 e^{-3x} \tag{2.3}$$

El valor de la resistencia mecánica es variable a lo largo de la membrana basilar de la forma:

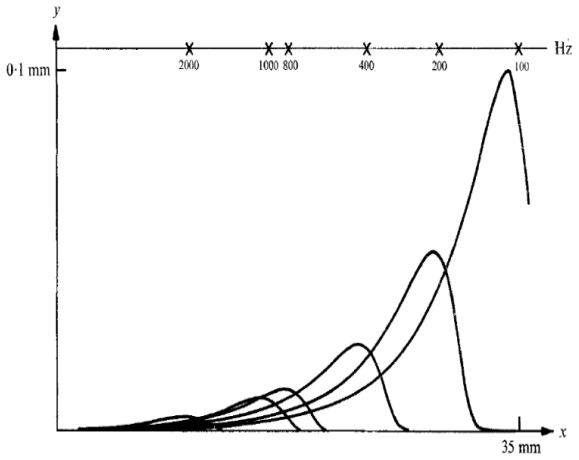


Figura 2.10: Respuesta de la membrana basilar por Lesser y Berkeley. (Original invertida) [Les72]

$$R_m(x) = 3000e^{-1,5x} \tag{2.4}$$

La masa a lo largo de la membrana basilar se considera constante con un valor de $m = 0{,}05$ en g/cm^2. Para la evaluación de su modelo Lesser y Berkeley utilizaron las frecuencias reportadas en el trabajo de Békésy de 100 Hz, 200 Hz, 400 Hz, 800 Hz, 1000 Hz y 2000 Hz. En la figura 2.10 se presenta la gráfica de sus resultados donde se muestra la amplitud de la envolvente de la serie de Fourier que emula el comportamiento de la membrana basilar, siendo el eje x la distancia a lo largo de la membrana basilar desde la ventana oval en mm y el eje y la amplitud de la envolvente en mm.

El modelo de la mecánica de fluidos en la cóclea de Lesser y Berkeley representa la mejor aproximación del comportamiento físico de la cóclea, sin embargo debido a que la solución de su modelo es una aproximación de la respuesta de la onda envolvente que se propaga sobre la membrana basilar, sus resultados obtenidos con respecto a las observaciones del comportamiento coclear resultaron diferentes, en años posteriores su modelo sirvió como base para que Allen obtuviera un modelo más aproximado del comportamiento coclear.

En 1971 Rhode realiza mediciones físicas del comportamiento de la cóclea, publicando sus resultados en un artículo para la JASA titulado *Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys using the Mössbauer Technique* [Rho71], tres años después en 1974 junto con Robles desarrolla experimentos para determinar el comportamiento no lineal de las vibraciones dentro de la cóclea, presentando sus resultados en un artículo para la JASA titulado *Evidence of Mossbauer experiments for nonlinear vibrarion in the cochlea* [Rho74], posteriormente Allen utiliza el modelo de Lesser y Berkeley para obtener los

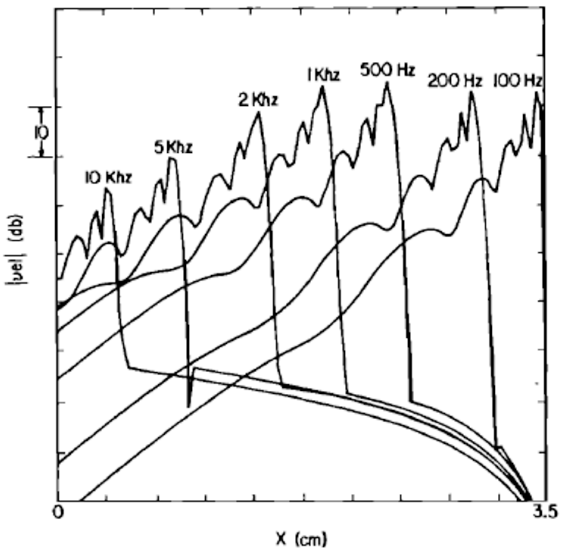


Figura 2.11: Respuesta de la membrana basilar por Allen. [All77]

parámetros de la membrana basilar utilizando la función de Green, publicando sus resultados para la JASA en 1977 en un artículo titulado *Two-dimensional cochlear fluid model: New Results* [All77]. En su modelo Allen considera que la constante de restitución varia a lo largo de la membrana basilar en $dyna/cm^2$ de la forma:

$$k(x) = 10^9 e^{-2ax} \tag{2.5}$$

Además que la resistencia mecánica varia a lo largo de la membrana basilar en $dyna \cdot seg/cm^2$ de la forma:

$$R_m = 300e^{-ax} \tag{2.6}$$

Siendo el valor de $a = 1{,}7$ en ambos casos, en este modelo la masa tiene un valor constante de $m = 0{,}1$ en g/cm^2. Para la evaluación de su modelo Allen utiliza las frecuencias de 100 Hz, 200 Hz, 500 Hz, 1000 Hz, 2000 Hz, 5000 Hz y 10000 Hz, en la figura 2.11 se muestra la gráfica de sus resultados, en el eje x se representa la distancia a lo largo de la membrana basilar en cm y en el eje y se respresenta la variación de velocidad del fluido dentro de la cóclea en dB, haciendo énfasis en que la variación de la velocidad es concordante con la variación en la amplitud de la membrana basilar en esas distancias específicas.

El trabajo de Allen modeló el comportamiento de la propagación de las ondas a lo largo de la membrana basilar en forma muy aproximada a los resultados experimentales obtenidos por Békésy y Rhode, sus estudios tienen la

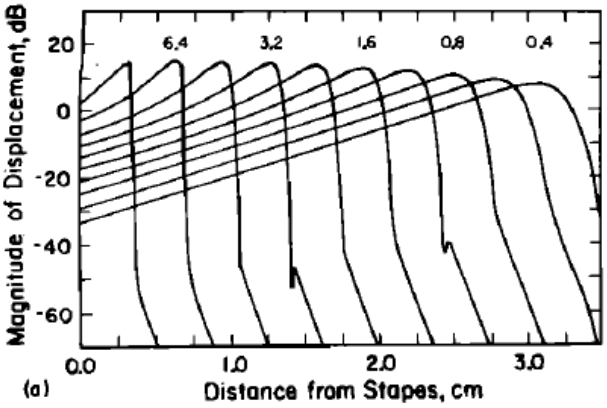


Figura 2.12: Respuesta de la membrana basilar por Neely. [Nee81]

ventaja de presentar una solución numérica para determinar los párametros de la membrana basilar.

Un trabajo posterior fue desarrollado en 1981 por Neely, en el cual se propone un modelo matemático en dos dimensiones de la cóclea y su solución numérica empleando aproximaciones por diferencias finitas usando la ecuación de Laplace, sus resultados fueron publicados para la JASA en un artículo titulado *Finite difference solution of a two-dimensional mathematical model of the cochlea* [Nee81]. Neely considera que la constante de elasticidad varía a lo largo de la membrana basilar en $dyna/cm^2$ de la forma:

$$k(x) = 10^9 e^{-2x} \tag{2.7}$$

En este modelo los valores de la resistencia mecánica y la masa por unidad de área a lo largo de la membrana basilar son constantes, teniendo sus valores respectivos de $R_m = 200\ dyna \cdot seg/cm^2$ y $m = 0{,}15\ g/cm^2$. En la figura 2.12 se presenta la gráfica de los resultados obtenidos por Neely para las frecuencias de prueba de 400 Hz, 570 Hz, 800 Hz, 1130 Hz, 1600 Hz, 2260 Hz, 3200 Hz, 4520 Hz, 6390 Hz y 9040 Hz, siendo el eje x la distancia a lo largo de la membrana basilar desde la ventana oval en cm y el eje y la magnitud de desplazamiento de la membrana basilar en dB correspondiente para cada frecuencia de prueba.

Cinco años más tarde en 1986 Neely y Kim amplian el modelo proponiendo elementos activos en la biomecánica de la cóclea, publicando sus resultados en un artículo para la JASA titulado *A model for active elements in cochlear biomechanics* [Nee86], siendo hasta estos momentos los parámetros más aproximados de la respuesta de la mecánica de la cóclea.

En años recientes se han utilizado estos parámetros para realizar estudios mas completos de la respuesta del comportamiento mecánico acústico de la cóclea, Elliott, Ku y Lineton en el año 2007 generan un modelo de espacio

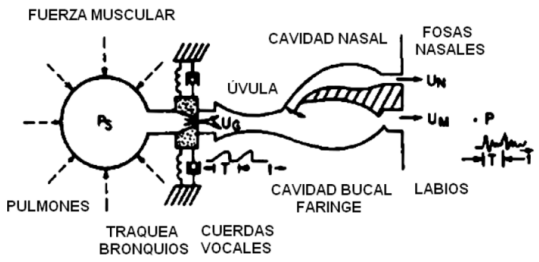


Figura 2.13: Elementos del aparato fonador.

estado de la mecánica de la cóclea utilizando estos parámetros, publicando sus resultados en un artículo para la JASA titulado *A state space model for cochlear mechanics* [Ell07], un año después en 2008 ellos mismos realizan estadísticas de estabilidad para su modelo y las publican en un artículo para la JASA titulado *Statistics of instabilities in a state space model of the human cochlea* [Ku08], posteriormente en 2011 Elliott, Lineton y Ni realizan un estudio del acoplamiento de fluidos en un modelo discreto de la mecánica de fluidos dentro de la cóclea, publicando sus resultados en un artículo para la JASA titulado *Fluid coupling in a discrete model of cochlear mechanics* [Ell11].

2.3. Fundamentos del procesamiento de voz

Fisiológicamente el proceso de generación de voz requiere de una fuente de energía constituida por el diafragma, los pulmones, los bronquios y la tráquea, con lo cual se realiza la espiración del aire. El aire pasa por un sistema vibratorio compuesto por las cuerdas vocales y la laringe, al pasar el aire a través de las cuerdas vocales las hace vibrar con una excitación pulsante a una frecuencia específica, la cual se denomina frecuencia fundamental o pitch. A continuación el flujo de aire pasa por una etapa de filtrado formada por la cavidad bucal y la cavidad nasal, dependiendo de cómo se encuentren articulados sus órganos se forma una caja de resonancia distinta, la cual genera un conjunto de frecuencias y atenúa el resto, lo cual da el timbre característico de voz para cada individuo [Rab78] [Qui99], en la figura 2.13 se muestran los elementos que conforman el aparato fonador.

La voz permite transmitir información por medio de señales acústicas en un intervalo de frecuencias entre 50 Hz y 8 kHz, este proceso de comunicación comienza con un mensaje generado en el cerebro, el cual es codificado mediante una concatenación de fonemas, posteriormente el cerebro genera las señales nerviosas que activan los componentes fisiológicos que generan acústicamente la voz, produciendo una onda. Esta onda acústica posteriormente llega al sistema auditivo del individuo receptor, en donde se convierte la energía mecánica en biopotenciales eléctricos para su interpretación en el cerebro [Rab81] [Del00],

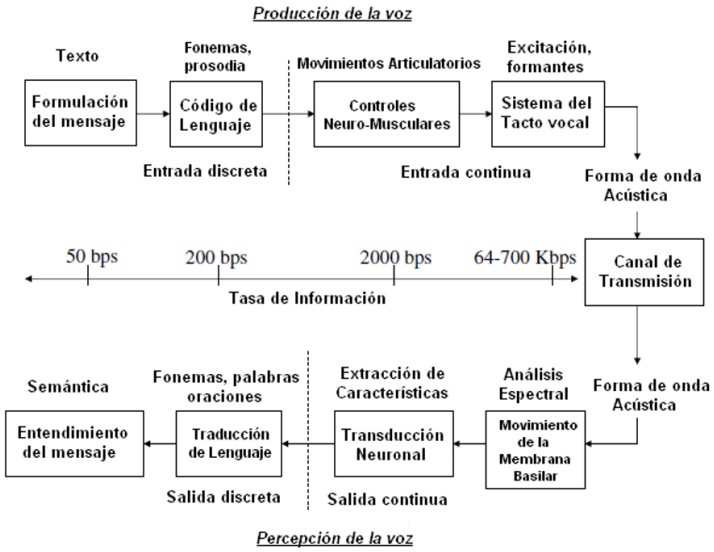


Figura 2.14: Proceso de comunicación por voz.

en la figura 2.14 se muestra el diagrama a bloques del proceso de comunicación por voz.

Las primeras investigaciones formales que se desarrollaron para entender la comunicación por voz fueron realizadas para el estudio de la comunicación telefónica, el primer antecedente es un artículo publicado para la AIEE por Nyquist en 1928 titulado *Certain Topics in Telegraph Transmision Theory* en el cual desarrolla la teoría matemática de la comunicación por voz [Nyq28], posteriormente Dudley publica un artículo para la JASA en 1939 titulado *Remaking Speech* en donde se plantea por primera vez los principios básicos de la codificación de voz [Dud39], sin embargo es hasta 1948 cuando Shanon publica en *The Bell System Technical Journal* su artículo titulado *A Mathematical Theory of Communication* en donde formaliza los procesos modernos de la teoría de comunicación por voz [Sha48].

Una vez desarrolladas las herramientas matemáticas de comunicación por voz, los investigadores se enfocaron en el proceso de producción de voz para entender su generación, el primer artículo que hace referencia al estudio de los fonemas vocálicos fué escrito por Peterson y Barney en 1952 para la JASA titulado *Control Methods Used in a Study of the Vowels* [Pet52], las características acústicas de la producción de las consonantes fueron estudiadas por Delattre, Liberman y Cooper en 1955 publicando sus resultados en un artículo para la JASA titulado *Acoustic Loci and Transitional Cues for Consonants* [Del55], posteriormente las consonantes fricativas fueron estudiadas en profundidad en 1961 por Heinz y Stevens quienes publicaron sus resultados en un artículo para

la JASA titulado *On the Propierties of Voiceless Fricative Consonants* [Hei61], un año después en 1962 Fujimura estudia las consonantes nasales y publica en la JASA su artículo titulado *Analysis of Nasal Consonants* [Fuj62].

La siguiente etapa de investigación una vez entendida la generación de la voz, fué modelar en forma acústica al sistema fónador, en 1961 Dunn presenta los métodos de medición de las formantes en las vocales en un artículo para la JASA titulado *Methods of Measuring Vowel Formant Bandwidths* [Dun61], en 1971 Sondhi y Gopinath presentan un estudio de la forma del tracto vocal considerando su respuesta al impulso a partir de los labios en un artículo para la JASA titulado *Determination of Vocal-Tract Shape from Impulse Response at the lips* [Son71], ese mismo año Rosenberg estudia la calidad de las vocales respecto a la variación de la forma del pulso glotal y lo publica en un artículo para la JASA titulado *Effect of Glottal Pulse Shape on the Quality of Natural Vowels* [Ros71], en un estudio posterior en 1974 Sondhi propone un modelo acústico de la propagación de ondas en el tracto vocal y lo expone en un artículo para la JASA titulado *Model for wave propagation in a lossy vocal tract* [Son74].

El modelo acústico de la producción de la voz describe en forma física al sistema fisiológico fonador, considera que la fuente de sonido es la que genera la energía acústica de la voz y es la excitación que se aplica al sistema, dependiendo de los sonidos a generar ésta puede tomar dos formas, una para los vocálicos y otra para los no vocálicos. Los sonidos vocálicos son producidos por las vibraciones de las cuerdas vocales contenidas en la laringe, cuando la glotis se cierra y libera la presión de aire, un efecto de Bernoulli se produce debido al incremento de velocidad lo cual hace que se decremente la presión entre las cuerdas generando un tren de impulsos cuyo periodo de vibración es el pitch, acústicamente el tracto vocal filtra la fuente de sonido y permite que algunas frecuencias se refuercen mientras otras se atenúen [Qua02] [Che88].

Cuando se obtuvo el modelo acústico de la generación de la voz, los investigadores estudiaron el problema de modelar en forma digital al sistema fonador, logrando con esto hacer un análisis discreto de la señal de voz y obtener los fundamentos para su síntesis, la excitación del tracto vocal en forma discreta fue estudiada por primera vez por Flanagan y Landgraf sus resultados fueron expuestos en un artículo publicado en 1968 para la IEEE titulado *Self-Oscillating Source for Vocal-Tract Synthesizers* [Fla68], un año mas tarde en 1969 Flanagan y Cherry publican para la JASA un artículo titulado *Excitation of Vocal-Tract Synthesizers* [Fla69], posteriormente en 1971 Atal y Hanauer presentan un modelo digital de producción de la voz en un artículo para la JASA titulado *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave* [Ata71], dos años más tarde en 1973 Portnoff realiza su tesis de Maestría en Ciencias en el Instituto Tecnológico de Massachusetts, desarrollando el modelo matemático para la simulación digital del tracto vocal y junto con Schafer presenta sus resultados en un congreso para la JASA en una ponencia titulada *Mathematical Considerations in Digital Simulations of the Vocalt Tract* [Por73].

El modelo digital de la producción de voz está basado en el modelo acústico de los elementos fisiológicos que generan la voz, se considera que el sistema es lineal y que el modelado de la fuente y del filtro del tracto vocal es separable. La

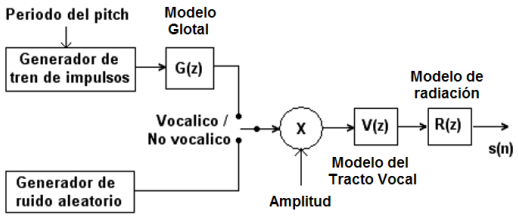


Figura 2.15: Modelo digital de producción de voz.

excitación que se aplica al sistema produce los sonidos de voz, éstos pueden ser sonidos vocálicos o no vocálicos, en el caso de los sonidos vocálicos es necesario un pulso glotal de excitación en forma de un tren de impulsos separados por el valor del periodo del pitch, el tren de impulsos es filtrado por un modelo del pulso glotal y multiplicado por un control de amplitud. Para los sonidos no vocálicos es necesaria una turbulencia de excitación ruidosa, la cual es provista por un generador de ruido gaussiano con control de amplitud, en algunos casos para generar algunos fonemas se utiliza la mezcla de los dos tipos de excitación. El modelo completo se ilustra en la figura 2.15, para formar los diferentes sonidos se tiene un interruptor que determina el modo de excitación [Che88] [Cas89].

La aplicación principal de los resultados del modelo digital de producción de voz fué el desarrollo de metodologías para síntesis de voz, en 1968 Rabiner estudia las técnicas para los sintetizadores de voz y publica sus resultados en dos artículos, uno para la JASA titulado *Digital-Formant Synthesizer for Speech-Synthesis Studies* [Rab68] y otro en colaboración con Gold para la IEEE titulado *Analysis of Digital and Analog Formant Synthesizers* [Gol68], un estudio sobre las señales de entrada en los sintetizadores es publicado por Winham y Steiglitz en 1970 en un artículo para la JASA titulado *Input Generators for Digital Sound Synthesis* [Win70], posteriormente ese mismo año Flanagan, Coker, Rabiner, Schafer y Umeda hacen una recopilación completa de los principios de la síntesis de voz y publican sus resultados en un artículo para la IEEE titulado *Synthetic Voices for Computers* [Fla70].

Con el advenimiento en los años setentas de las computadoras digitales se propició que los sistemas de procesamiento de señales analógicos fueran implementados en forma discreta, en 1965 Cooley y Tukey desarrollan los fundamentos del análisis de Fourier en forma discreta en un artículo para *Math Computation* titulado *An Algorithm for the Machine Calculation of Complex Fourier Series*, posteriormente en 1967 Helms implementa los filtros digitales y la transformada de Fourier en forma discreta, sus estudios fueron publicados en un artículo de la IEEE titulado *Fast Fourier Transform Method of Computing Difference Equations and Simulating Filters* [Hel67], en los comienzos de los años setentas Schafer y Rabiner aplican las técnicas desarrolladas en la década anterior al análisis de voz, sus resultados son publicados en un artículo en 1975 para la IEEE titulado *Digital Representations of Speech Signals* [Sch75],

un año después en 1976 publican un segundo artículo también para la IEEE titulado *Digital Techniques for Computer Voice Response: Implementations and Applications* [Rab76], en estos dos artículos se presentan los fundamentos de las técnicas actuales del análisis de la señal de voz.

2.4. Parametrización de la señal de voz

Debido a que la representación de la señal de voz en el dominio del tiempo y en el dominio de la frecuencia implica el uso de ventanas de análisis grandes, las investigaciones posteriores se enfocaron en poder representar la señal de voz con un número reducido de valores denominados coeficientes, lo cual condujo el desarrollo de varias metodologías para su obtención, las primeras utilizaron el modelo digital de producción de voz, posteriormente se emplearon modelos basados en la audición que proporcionaron mejoras notables al usarlos en procesos de reconocimiento de voz.

La metodología del cepstrum fue desarrollada por Bogert, Healy y Tukey en 1963 y presentada en el *Symposium on Time Series Analysis* en una ponencia titulada *The Quefrency Analysis of Time Series for Echoes* [Bog63], un año después Noll utiliza esta técnica para el procesamiento de voz y publica sus resultados para la JASA en dos artículos, el primero en 1964 titulado *Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection* [Nol64] y el segundo en 1967 titulado *Cepstrum Pitch Determination* [Nol67], con esta metodología se obtienen los coeficientes cepstrales.

El cepstrum $c(\tau)$ se define como la transformada de Fourier inversa de la amplitud espectral logarítmica en tiempo corto $|X(\omega)|$, permite representar en forma separada la envolvente espectral de la señal de voz y su estructura fina. Esta metodología está basada en considerar a la señal de voz $x(t)$ como la respuesta del filtro equivalente de la articulación del tracto vocal, el cual es excitado por una fuente cuasiperiódica $g(t)$, por lo tanto $x(t)$ está dada por la convolución de $g(t)$ y la respuesta al impulso del tracto vocal $h(t)$ [Fur01], estando expresado de la siguiente forma:

$$x(t) = \int_0^t g(\tau)h(t-\tau)d\tau \tag{2.8}$$

Si se considera que $X(\omega)$, $G(\omega)$ y $H(\omega)$ son las trasformadas de Fourier de $x(t)$, $g(t)$ y $h(t)$ respectivamente, la ecuación anterior se puede expresar de la forma:

$$X(\omega) = G(\omega)H(\omega) \tag{2.9}$$

Si $g(t)$ es una función periódica $|X(\omega)|$ está representada por la línea espectral, y cuando $|X(\omega)|$ se calcula con la transformada de Fourier de una ventana de análisis de la señal de voz la forma de los picos que se observan tienen iguales intervalos a lo largo del eje de la frecuencia, el logaritmo de la función $|X(\omega)|$ se define como:

$$log|X(\omega)| = log|G(\omega)| + log|H(\omega)| \tag{2.10}$$

Para obtener el cepstrum se aplica la transformada de Fourier inversa al logaritmo de $|X(\omega)|$, quedando la siguiente expresión:

$$c(\tau) = F^{-1}log|X(\omega)| = F^{-1}log|G(\omega)| + F^{-1}log|H(\omega)| \tag{2.11}$$

Donde F es la transformada de Fourier de los terminos del lado derecho de la ecuación, el primero corresponde a la estructura espectral fina y el segundo a la envolvente espectral. Cuando se utiliza la transformada de Fourier discreta para calcular el cepstrum, es necesario colocar un valor base N grande de tal manera que elimine el aliasing producido durante el muestreo de la voz, lo cual se expresa de la forma:

$$c_n = \frac{1}{N}\sum_{k=0}^{N-1} log|X(k)|e^{j2\pi kn/N} \qquad 0 \leq n \leq N-1 \tag{2.12}$$

En 1968 se desarrolla la metodología de predicción lineal tomando como base al modelo digital de producción de voz, fué implementada por Atal y Schroeder y en forma paralela por Itakura y Saito, fué publicada en 1971 por Atal y Hanauer en un artículo para la JASA titulado *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave* [Ata71], empleando este método se obtienen los coeficientes LPC (Linear Prediction Coeficients).

El metódo de predicción lineal considera que una muestra de voz $s(n)$ es predecida a partir de la suma lineal de las muestras previas afectadas por un coeficiente a_i, lo cual se expresa de la forma:

$$\widehat{s}(n) = a_1 s(n-1) + a_2 s(n-2) + ... + a_p s(n-p) \tag{2.13}$$

Donde el conjunto de $\{a_k\}$ son los coeficientes de predicción lineal determinados por la minimización de la diferencia cuadrática media entre la muestra de voz actual y la precedida linealmente. La ecuación de predicción lineal puede ser escrita también de la forma:

$$\widehat{s} = \sum_{k=1}^{p} a_k s(n-k) \tag{2.14}$$

A partir del modelo digital de producción de voz se establece que la función de transferencia del filtro digital $H(z)$ produce la salida de voz $s(n)$ la cual está dada por la excitación $u(n)$, ya sea un tren de impulsos o un ruido aleatorio, estando esta función representada de la forma:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{G}{A(z)} \tag{2.15}$$

Donde G es la ganancia aplicada al filtro, siendo éste el modelo todos polos de la producción de la voz, donde las raíces del polinomio del denominador $A(z)$

son los polos del sistema. En el dominio del tiempo la ecuación del modelo de producción de la voz se expresa en una forma diferencial como:

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{2.16}$$

Se puede notar la forma similar a la ecuación de predicción lineal, si los coeficientes del predictor son iguales a los coeficientes del filtro, la ecuación de predicción lineal se aproxima muy bien al modelo de la voz. Si se considera la forma de excitación $A(z)$ para generar una secuencia de voz usando la forma del filtro inverso, el sistema se puede expresar de la siguiente forma:

$$S(z)A(z) = U(z)H(z)A(z) = U(z)G \tag{2.17}$$

Lo cual da una secuencia llamada residuo que se aproxima a la excitación aplicada, siendo necesario encontrar un conjunto de valores $\{a_k\}$ tales que $A(z)$ sea el inverso de $H(z)$. Esto se hace utilizando la ecuación para encontrar una estimación $\widehat{s}(n)$ de las muestras de la señal de voz $s(n)$, estando este error dado por:

$$e(n) = s(n) - \widehat{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{2.18}$$

Para obtener los coeficientes del predictor se minimiza el error de predicción cuadrático medio en un segmento corto de la señal de voz, encontrando la secuencia a_k para un intervalo especificado m, utilizando la expresión:

$$E_n = \sum_{m} |e_n(m)|^2 \tag{2.19}$$

Siguiendo la aproximación de mínimos cuadrados de considerar $\partial E_n/\partial a_i = 0$, $i = 1, 2, ..., p$, se obtiene el siguiente conjunto de p ecuaciones:

$$\sum_{k-1}^{p} a_k \sum_{m} (m-i)s_n(m-k) = \sum_{m} s_n(m-i)s_n(m) \qquad i = 1, 2, ..., p \tag{2.20}$$

Empleando el método de autocorrelación se asume que $-\infty \leq m \leq \infty$ y que la señal de voz está ventaneada tal que es cero afuera del intervalo de interés, siendo $s_n(m) = s(m+n)\omega(n)$, haciendo que la ecuación anterior se convierta en:

$$\sum_{k-1}^{p} a_k R_n(|i-k|) = R_n(i) \qquad i = 1, 2, ..., p \tag{2.21}$$

Donde la matriz resultante del lado izquierdo de la ecuación es simétrica, positiva y cumple con las condiciones de Toeplitz, la solución de este sistema de ecuaciones proporciona los coeficientes LPC a partir de la condición:

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m)s_n(m+k) \tag{2.22}$$

Posteriormente se propone la metodología del análisis homomórfico tomando como base el cepstrum, fué desarrollada por Oppenheim y Schafer en 1968 y presentada en un artículo para la IEEE titulado *Homomorphic Analysis of Speech* [Opp68], posteriormente en 1969 un año después Oppenheim publicaría un segundo artículo para la JASA titulado *A Speech Analysis-Synthesis System Based on Homomorphic Filtering* [Opp69], con esta metodología se obtienen los parametros CLPC (Cepstrals Lineal Prediction Coeficients).

Los CLPC se basan en la descomposición no lineal del sistema en factores independientes considerando un caso especial del cepstrum en el cual $X(\omega) = H(z)$, donde $H(z)$ es la transformada z de la respuesta al impulso de una estimación del modelo todos polos de producción de voz por predicción lineal, lo cual se expresa de la forma:

$$H(z) = \frac{1}{1 + \sum_{i=1}^{p} \alpha_i z^{-i}} \tag{2.23}$$

Lo anterior implica que el espectro todo polos $H(z)$ es usado para la densidad espectral de la señal de voz, complementándose con la expansión del cepstrum en una forma compleja al reemplazar la transformada de Fourier Discreta, la transformación logarítmica y la transformada discreta inversa de Fourier. El cepstrum complejo para una secuencia $x(n)$ es representado por $\widehat{c}_n$ y sus transformadas z son $X(z)$ y $C(z)$, el sistema es descrito por:

$$\widehat{C}(z) = log[X(z)] \tag{2.24}$$

Si se diferencian ambas partes de la ecuación por z^{-1} y se multiplican por $X(z)$ se obtiene la siguiente expresión:

$$X(z)\widehat{C}'(z) = X'(z) \tag{2.25}$$

Esta ecuación permite emplear un sistema de ecuaciones recursivas con lo cual se pueden obterner los coeficientes CLPC.

$$\widehat{c}_1 = -\alpha_1$$

$$\widehat{c}_n = -\alpha_n - \sum_{m=1}^{n-1}(1 - \frac{m}{n})\alpha_m \widehat{c}_{n-m} \qquad 1 < n \leq p \tag{2.26}$$

$$\widehat{c}_n = -\sum_{m=1}^{p}(1 - \frac{m}{n})\alpha_m \widehat{c}_{n-m} \qquad p < n$$

En los años posteriores la investigación se condujo a utilizar los modelos perceptuales de la voz, destacando el trabajo de Davis y Mermelstein realizado en

1980 basado en la respuesta auditiva de la escala de Mel, sus resultados fueron publicados en un artículo para la IEEE titulado *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences* [Dav80], empleando esta metodología se obtienen los parámetros MFCC (Mel Frequency Cepstrals Coeficients).

Esta metodología realiza una representación similar al cepstrum tomando como base una escala de frecuencias no lineal que se aproxima al comportamiento del sistema auditivo humano conocida como escala de Mel [Ros07], la cual está definida como:

$$\beta(f) = 1125ln(1 + \frac{f}{700}) \tag{2.27}$$

Dada una transformada discreta de la señal de entrada, se distribuye el comportamiento espectral en bandas de frecuencias, mediante la implementación de un banco de filtros espaciados entre dos puntos límites respecto a la escala de Mel, estableciendo una frecuencia límite inferior f_l y una frecuencia límite superior f_h lo anterior queda expresado de la siguiente forma:

$$f(m) = \frac{N}{F_s}\beta^{-1}(\beta(f_h) + m\frac{\beta(f_l) - \beta(f_h)}{M+1}) \tag{2.28}$$

Donde β^{-1} está dada por:

$$\beta^{-1}(b) = 700(e^{\frac{b}{1125}} - 1) \tag{2.29}$$

Se construye un arreglo de filtros triangulares en un intervalo definido desde una frecuencia de corte hasta una frecuencia menor al criterio de Nyquist, a partir de la expresión:

$$H_m(k) = \begin{cases} 0 & k < f[m-1] \\ \frac{2(k-f[m-1])}{(f[m+1]-f[m-1])(f[m]-f[m-1])} & f[m-1] \leq k \leq f[m] \\ \frac{2(f[m+1]-k)}{(f[m+1]-f[m-1])(f[m]-f[m-1])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \tag{2.30}$$

A continuación se procede a calcular el logaritmo de la energía de cada filtro, mediante la siguiente ecuación:

$$S(m) = ln(\sum_{k=0}^{N-1} |X(k)|^2 H_m(k)) \quad 0 \leq m \leq M \tag{2.31}$$

Los coeficientes cepstrales en frecuencia en escala de Mel MFCC se obtienen al aplicar la transformada discreta coseno en las salidas de cada filtro M, lo cual está definido de la forma:

$$c(n) = \sum_{m=0}^{M-1} S(m)cos(\pi n(\frac{m - \frac{1}{2}}{M})) \quad 0 \leq n \leq N-1 \tag{2.32}$$

Otro trabajo basado en los modelos de la audición fué realizado en 1990 por Hermansky utilizando la predicción lineal perceptual, sus resultados fueron publicados en un artículo para la JASA titulado *Perceptual linear predictive (PLP) analysis of speech* [Her90], a partir de esta metodología se obtienen los parámetros PLP.

Consiste en obtener la transformada de Fourier de un segmento corto de voz y elevar al cuadrado la parte real y la parte imaginaria, obteniendo el espectro en potencia de la siguiente forma:

$$P(\omega) = Re[S(\omega)]^2 + Im[S(\omega)]^2 \tag{2.33}$$

El espectro de potencia $P(\omega)$ es distorcionado a lo largo del eje ω a partir de la respuesta en frecuencia Ω basada en la escala de Bark, la cual representa la respuesta de enmascaramiento al ruido del sistema auditivo y está definida de la forma:

$$Bark(f) = 6ln\{(f/600) + \sqrt{(f/600)^2 + 1}\} \tag{2.34}$$

Con base en la escala de Bark el espectro distorcionado se expresa de la forma:

$$\Omega(\omega) = 6ln\{\frac{\omega}{1200\pi} + [(\omega/1200\pi)^2 + 1]^{0,5}\} \tag{2.35}$$

El espectro en potencia distorcionado resultante es entonces convolucionado con el espectro de potencia de la curva de banda crítica de enmascaramiento $\Psi(\Omega)$ dada por:

$$\Psi(\Omega) \begin{cases} 0 & \Omega < -1{,}3 \\ 10^{2,5(\Omega+0,5)} & -1{,}3 \leq \Omega \leq -0{,}5 \\ 1 & -0{,}5 < \Omega < 0{,}5 \\ 10^{-1(\Omega-0,5)} & 0{,}5 << \Omega << 2{,}5 \\ 0 & \Omega > 2{,}5 \end{cases} \tag{2.36}$$

La convolución discreta de $\Psi(\Omega)$ otorga las muestras del espectro de potencia en la banda crítica, lo cual se expresa como:

$$\theta(\Omega_i) = \sum_{\Omega=-1,3}^{2,5} P(\Omega - \Omega_i)\Psi(\Omega) \tag{2.37}$$

Los coeficientes PLP se obtienen al seleccionar un número entero de muestras espectrales que cubren la banda completa de análisis.

2.5. Metodologías de reconocimiento de voz

Una vez que se obtuvo un arreglo finito de valores característicos que representan un segmento de la señal de voz, se procedió a investigar un conjunto de técnicas para realizar su reconocimiento las cuales se fundamentaron en cuatro

paradigmas, el primero basado en templates, el segundo en el conocimiento, el tercero en los procesos estocásticos y el cuarto en los modelos conexionistas [Wai90]. Sin embargo debido a que los primeros métodos matemáticos desarrollados para la comunicación por voz condujeron a proponer la implementación de máquinas que entendieran mensajes hablados, los dos primeros artículos que mostraron las espectativas para este tipo de trabajo conservaron este primer planteamiento, ambos fueron publicados para la IEEE en 1976, el primero por Flanagan titulado *Computers that Talk and Listen: Man-Machine Communication by Voice* [Fla76] y el segundo por Reddy titulado *Speech Recognition by Machine: A review* [Red76].

El primer paradigma considera que las unidades de voz son representadas por medio de templates de la misma forma como la señal de voz se encuentra en el dominio del tiempo, y por lo tanto métricas de distancia o similitud se utilizan para comparar los templates con el objetivo de encontrar la mejor relación, se utiliza la programación dinámica para resolver el problema de la variabilidad temporal [Wai90] [Ben08]. La primera metodología desarrollada para el reconocimiento de voz utilizando este paradigma fué el alineamiento temporal dinámico (DTW) y se basó en la comparación de patrones, fué propuesta por Itakura en 1975 y publicada en un artículo para la IEEE titulado *Minimum Prediction Residual Principle Applied to Speech Recognition* [Ita75], posteriormente tres años después en 1978 Sakoe y Chiba optimizan esta metodología utilizando la programación dinámica, publican su propuesta en un artículo para la IEEE titulado *Dynamic Programming Algorithm Optimization for Spoken Word Recognition* [Sak78], durante el siguiente año en 1979 se realiza un trabajo para el reconocimiento de palabras aisladas usando DTW por Rabiner, Levinson, Rosenberg y Wilpon enfocándose al reconocimiento del locutor, sus resultados se publican en un artículo para la IEEE titulado *Speaker - Independent Recognition of Isolated Words Using Clustering Techniques* [Rab79], por su parte Sakoe estudia la aplicación del DTW al reconocimiento de palabras conectadas publicando un artículo para la IEEE titulado *Two-Level DP-Matching-A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition* [Sak79], posteriormente en 1981 Rabiner emplea en un mismo trabajo el reconocimiento de palabras aisladas y conectadas, publicando sus resultados en un artículo para la IEEE titulado *Isolated and Connected Word Recognition - Theory and Selected Applications* [Rab81].

La comparación de patrones empleando DTW presentó buenos resultados para aplicaciones no robustas o de vocabulario limitado, debido a esto las investigaciones posteriores se enforcaron en los métodos de minimización de distancias métricas buscando mejorar los resultados obtenidos hasta entonces y en 1980 Linde, Buzo y Gray desarrollan la técnica de Cuantificación Vectorial (VQ) logrando una alta eficiencia, sus resultados se publican en un artículo para la IEEE titulado *An Algorithm for Vector Quantizer Design* [Lin80], tres años después en 1983 Gersho y Cuperman realizan estudios sobre codificación utilizando VQ, publicando sus resultados en un artículo para la IEEE titulado *Vector Quantization: A Pattern-Matching Technique for Speech Coding* [Ger83], posteriormente en 1984 Gray sintetiza la VQ en un artículo para la IEEE titulado

Vector Quantization [Gra84].

Las metodologías anteriores demostraron que el paradigma basado en templates ha sido satisfactorio particularmente para aplicaciones simples que requieren un mínimo de caracterésticas, una crítica a las metodologías basadas en templates es que no facilitan el uso del conocimiento en el proceso del reconocimiento de voz, otra desventaja de este paradigma es su limitada capacidad para generalizar. Un enfoque de estudio diferente es el paradigma basado en conocimiento, el cual emula el conocimiento de la voz humana utilizando sistemas expertos, sin embargo los sistemas basados en estas reglas han tenido éxito limitado. Una alternativa más exitosa consiste en segregar el conocimiento de algoritmos e integrar conocimiento dentro de otros paradigmas de índole matemático, la adición del conocimiento tendió a encontrar otros enfoques de manera sustancial [Wai90].

El paradigma estocástico, es en algunos aspectos, similar al basado en templates, la diferencia más grande es que utiliza modelos probabilísticos debido a que la señal de voz puede ser caracterizada como un proceso aleatorio paramétrico y que estos parámetros pueden ser determinados empleando modelos probabilísticos. Para su implementación se utilizan típicamente modelos ocultos de Markov (HMM) los cuales pueden modelar la incertidumbre inherente en el reconocimiento de voz, con lo cual resuelven de manera simultánea la segmentación y el problema de clasificación, lo cual los hace particularmente adecuados para el reconocimiento de voz contínua, los sistemas de gran escala utilizan este paradigma [Rab93].

Una característica adicional de los HMM es que pueden realizar ciertas consideraciones acerca de la estructura del lenguaje dentro del proceso de reconocimiento de voz, y realizar la estimación de los parámetros del sistema por medio de las estructuras que son correctas. Esto tiene la ventaja de reducir el problema de aprendizaje, pero la desventaja de tratar con consideraciones regularmente incorrectas [Fur01].

El primer estudio con este enfoque se realizó en 1975 por Baker sus resultados fueron publicados para la IEEE en un artículo titulado *The DRAGON System-An Overview* [Bak75], posteriormente a principios de los años ochenta Bahl, Jelinek y Mercer aplican esta metodología para el reconocimiento de voz continua, sus resultados fueron publicados en 1983 en un artículo para la IEEE titulado *A Maximum Likelihood Approach to Continuous Speech Recognition* [Bah83], un trabajo posterior fue desarrollado por Roucos y Dunham para el reconocimiento de voz continua basado en fonemas usando modelos estocásticos, sus resultados fueron publicados en 1987 en un artículo para la IEEE titulado *A Stochastic Segment Model for Phoneme-Based Continuous Speech Recognition* [Rou87], dos años mas tarde Rabiner, Wilpon y Soong emplearon los HMM para el reconocimiento eficiente de dígitos conectados, publicando sus resultados en 1989 en un artículo para la IEEE titulado *High Performance Connected Digit Recognition Using Hidden Markov Models* [Rab89a], ese mismo año Rabiner publica una síntesis del empleo de los HMM aplicados al reconocimiento de voz en un artículo para la IEEE titulado *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition* [Rab89b].

El paradigma conexionista difiere de los HMM en que muchas de estas consideraciones necesarias no son requeridas. Los paradigmas conexionistas utilizan la representación distribuida de muchos nodos simples, en donde las conexiones son entrenadas para reconocer la voz siendo las redes neuronales (NN) los modelos más utilizados con este tipo de enfoque, este paradigma representa el desarrollo más reciente en reconocimiento de voz [Hil00] [Mar07].

El desarrollo de los modelos de las redes neuronales (NN) en las técnicas de inteligencia artificial condujo a investigaciones posteriores para su implementación en los sistemas de reconocimiento de voz, tomando como fundamento el paradigma conexionista y su aproximación a los modelos desarrollados del sistema fisiológico cerebral en el funcionamiento de las neuronas. El primer trabajo empleando estas técnicas fue desarrollado por Kohonen en 1988, sus resultados fueron publicados para la IEEE en un artículo titulado *The "Neural" Phonetic Typewriter* [Koh88], posteriormente en 1989 se desarrollan diferentes proyectos basados en NN para el reconocimiento de voz y se publican para la IEEE, Waibel, Hanazawa, Hinton, Shikano y Lang estudiaron el reconocimiento de fonemas usando NN con retraso en tiempo, presentaron sus resultados en un artículo titulado *Phoneme Recognition Using Time-Delay Neural Networks* [Wai89a], en forma paralela Waibel, Sawai y Shikano estudian el reconocimiento de consonantes por construcción modular, presentando sus resultados en el artículo *Consonant Recognition by Modular Construction of Large Phonemic Time-Delay Neural Networks* [Wai89b].

Una vez que se desarrollaron las diferentes metodologías para el reconocimiento de voz, las siguientes investigaciones plantearon utilizar sistemas híbridos entre estas metodologías buscando un mejor desempeño, el primer proyecto de sistemas híbridos fué propuesto en 1989 por Sakoe, Isotani, Yoshida, Iso y Watanabe, mostró como utilizar las NN para el reconocimiento de palabras independientes del locutor utilizando la programación dinámica, presentando sus resultados para la IEEE en un artículo titulado *Speaker-Independent Word Recognition Using Dynamic Programming Neural Networks* [Sak89], posteriormente en 1990 Iso y Watanabe desarrollaron un modelo de predicción neuronal para el reconocimiento de palabras independientes del locutor, mostrando sus resultados en un artículo para la IEEE titulado *Speaker-Independent Word Recognition using a Neural Prediction Model* [Iso90], un año después McDermott y Katagiri, desarrollan un sistema híbrido entre la VQ y los HMM, mostrando sus resultados en un artículo para la IEEE titulado *LVQ-Based Shift-Tolerant Phoneme Recognition* [McD91], durante el comienzo de la década de los noventa hasta los años actuales los trabajos siguientes emplearon este tipo de sistemas híbridos.

Algunos de los problemas continúan siendo un reto, como la reducción del tiempo de reconocimiento y un modelado de las condiciones secuenciales. Las oraciones siempre contienen ambigüedades que no pueden ser resueltas por el reconocimiento acústico fonético a nivel de palabras. El reconocimiento de oraciones satisfactorias debe de alguna manera incorporar condiciones que trasciendan este nivel, incluyendo condiciones de semántica, prosodia y sintaxis. Para lo cual el procesamiento del lenguaje realiza estas tareas. De tal manera que

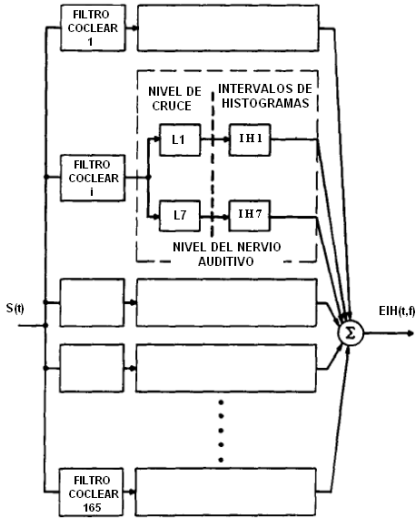


Figura 2.16: Modelo EIH.

la mayoría de los sistemas de reconocimiento a gran escala realizan el reconocimiento satisfactorio tomando en consideración la potencia de los modelos del lenguaje [Qui99]. En el apéndice dos se presentan los fundamentos matemáticos de las metodologías de reconocimiento de voz descritas en esta sección.

2.6. Análisis de voz usando la respuesta del oído interno

Las investigaciones posteriores en la etapa de extracción de coeficientes, se basaron en el empleo de la respuesta fisiológica de los modelos del oído interno utilizando arreglos de bancos de filtros. El único trabajo reportado hasta el momento fué desarrollado por Ghitza a finales de los años ochenta y descrito en 1992 en el libro *Advances in Speech Signal Processing* [Fur92], está fundamentado en un modelo fenomenológico de los cilios desarrollado por Allen y publicado en 1985 para la IEEE en un artículo titulado *Cochlear Modelling* [All85], en su trabajo Ghitza considera que los cilios se comportan como un banco de filtros que emulan la selectividad en frecuencia a lo largo de la membrana basilar, teniendo a la salida de los filtros un patrón que simula la respuesta del nervio auditivo, por lo cual el modelo fué llamado histogramas ensamblados por intervalos (EIH), en la figura 2.16 se muestra el diagrama a bloques del sistema.

En el modelo EIH el movimiento mecánico de la membrana basilar se modela usando 165 filtros igualmente espaciados en una escala de frecuencia logarítmica entre 150 Hz y 7 kHz, las características para cada filtro se basan en las curvas

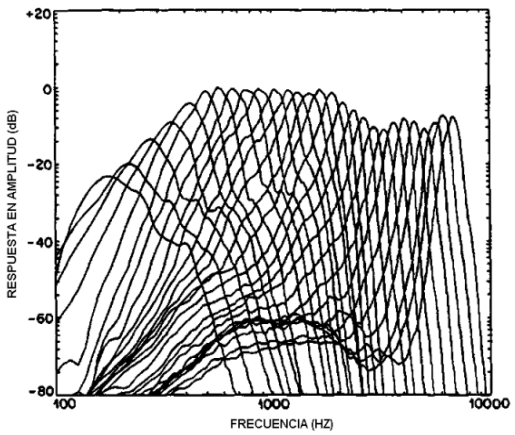


Figura 2.17: Respuesta en frecuencia del modelo EIH.

de la respuesta del sistema auditivo, considerando sus características de fase como mínimas y su medida de ganancia relativa a la frecuencia central refleja el valor correspondiente de la función de transferencia del oído medio, en la figura 2.17 se muestra la respuesta de amplitud contra frecuencia de 28 filtros correspondientes al intervalo entre 100 Hz y 10 kHz.

Una serie de bloques detectores de niveles de cruce modelan la actividad de los cilios de transducción de energía mecánica a biopotenciales eléctricos, los niveles de detección de cada bloque están distribuidos en forma aleatoria simulando la variabilidad de los diámetros de los filamentos de los cilios junto con sus conexiones sinápticas. La forma logarítmica de la respuesta en frecuencia representa una correlación entre los puntos de audición sobre la membrana basilar, cada canal IHC tiene siete detectores de nivel de cruce. Si la magnitud de salida del filtro es baja únicamente un nivel de cruce se habilita, para señales de magnitudes grandes varios niveles son habilitados creando un área de actividad específica. Los patrones de salida de los detectores de cruce representan la actividad del nervio auditivo, la cual a su vez es la entrada a la etapa central del procesamiento neuronal.

Conceptualmente el modelo EIH es una medida del incremento de la actividad neuronal a lo largo del nervio auditivo y matemáticamente es una función de densidad de probabilidad de condición reducida de los intervalos entre medidas sucesivas a lo largo de la simulación del nervio auditivo en una región tiempo frecuencia característica. Como una consecuencia de los detectores de cruce multinivel, la representación del modelo EIH preserva la información acerca de la energía total de la señal, si se considera el caso de una señal senoidal de entrada dada de la siguiente forma.

$$s(t) = Asen(2\pi f_0 t) \tag{2.38}$$

Donde f_0 es la frecuencia característica del canal selecionado y A es la intensidad, la salida del filtro coclear activará unicamente algunos detectores de cruce de nivel bajo. El histograma es escalado en unidades de frecuencia, estando los intervalos determinados a partir de f_0 para cada bloque, para la señal de entrada propuesta todos los intervalos son los mismos, resultando en un histograma en el cual la magnitud de cada bloque excepto el de f_0 son cero. Cuando la amplitud de la señal se incrementa, muchos niveles son activados contribuyendo a la activación del bloque de f_0, la respuesta en amplitud y fase de los filtros están dadas de la forma.

$$|H_i(f)| \tag{2.39}$$

$$\phi_i(f) \quad i = 1, 2, ..., 165 \tag{2.40}$$

Debido a la forma de los filtros más de un canal coclear contribuye al bloque de f_0, proporcionando que $A|H_i(f_0)|$ exceda cualquier umbral de los niveles de cruce, lo cual se expresa mediante la siguiente ecuación.

$$s_i(t) = A|H_i(f_0)|sen(2\pi f_0 t + \phi(f_0)) \tag{2.41}$$

Una de las metas del procesamiento de señales basado en el modelado del oído interno, es hacer el análisis de la señal más robusto al ruido mediante procedimientos de análisis espectrales alternativos basados en bancos de filtros.

2.7. Modelado mecánico acústico del oído interno en reconocimiento de voz

Los modelos del oído interno basados en la mecánica de fluidos descritos en el estado del arte se basan en diferentes métodos para dar solución al comportamiento mecánico acústico de la cóclea, siendo el que mejor se ajusta a las observaciones de Békésy el desarrollado por Lesser y Berkeley. Sin embargo estos modelos presentan deficiencias en la relación frecuencia distancia, ya sea en las bajas frecuencias cercanas al helicotrema como es el caso del modelo de Peterson o en las altas frecuencias cercanas al ápice como los modelos de Allen y Neely.

Para tratar de dar solución a esta problemática, en este trabajo de tesis se propone una nueva metodología usando análisis por resonancia en el modelo de la membrana basilar como un conjunto de osciladores armónicos forzados propuesto por Lesser y Berkeley. Con este planteamiento es posible determinar la amplitud máxima de desplazamiento sobre la membrana basilar para cada frecuencia de excitación, dependiendo solamente de las características físicas de masa, constante de elasticidad y resistencia mecánica a lo largo de ella. Siendo la principal ventaja del análisis por resonancia poder utilizar cualquier valor de frecuencia para obtener la distancia sobre la membrana basilar donde se

presenta la máxima amplitud, lo cual es concordante con la teoría de los puntos de audición de Békésy.

Debido a que las metodologías de parametrización de la voz basadas en la percepción auditiva han dado mejores resultados que las que utilizan los modelos de generación de la voz al ser implementadas en sistemas de reconocimiento, este trabajo de tesis propone emplear la solución del análisis por resonancia para hacer una aproximación de la forma en que la información es procesada por el oído interno, lo cual sólo ha sido propuesto por Ghitza para el análisis de señales de voz [Fur92]. Este nuevo planteamiento se basa en el desarrollo de un banco de filtros que modelan la respuesta del óido interno en forma similar a como lo hacen los filtros empleados por los MFCC, usando el análisis por resonancia y el método de diferencias finitas para determinar los intervalos de frecuencias de cada filtro. Posteriormente se hace su implementacion en HTK para evaluar su desempeño en pruebas de reconocimiento de voz.

Capítulo 3

Propuesta de solución

A vos os toca lo demás. Dios no quiere hacerlo todo para no quitarnos el libre albedrío ni la parte de gloria que nos corresponde...
Nicolas Maquiavelo. El príncipe, Florencia 1513.

El presente capítulo describe la propuesta de solución de este trabajo de Tesis al problema del modelado mecánico acústico del oído interno usando análisis por resonancia y su aplicación al problema de la parametrización de la señal de voz para su reconocimiento. Primero se presenta la mecánica de fluidos necesaria para la descripción del movimiento de la perilinfa y la endolinfa dentro de la cóclea. A continuación se describe la solución del comportamiento físico de la membrana basilar como un sistema de osciladores armónicos forzados concatenados junto con sus condiciones límites. Después se muestra en forma detallada la solución del modelo empleando análisis por resonancia hasta la obtención de un modelo frecuencia distancia. Por último se presenta una propuesta de metodología de parametrización de la voz mediante la implementación de un banco de filtros empleando diferencias finitas y la solución desarrollada.

3.1. Mecánica de fluidos en la cóclea

El movimiento de los fluidos en la cóclea puede ser descrito a partir de las ecuaciones del movimiento para un fluido incompresible y viscoso, considerando que la cóclea está dividida en dos compartimientos rectangulares separados por la membrana basilar y llenos de un fluido de características similares a la perilinfa y endolinfa. El compartimiento superior corresponde a la escala vestibular y el compartimiento inferior corresponde a la escala timpánica, por simplicidad se omite la escala media, en la figura 3.1 se muestra el diagrama del modelo de la mecánica de fluidos en la cóclea, mostrando las condiciones límites del desplazamiento de la membrana basilar [Kee08].

Se considera que el fluido contenido dentro de la cóclea es incompresible

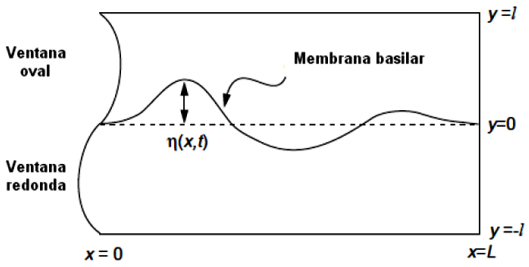


Figura 3.1: Mecánica de fluidos en la cóclea.

y presenta características homogéneas, por lo cual se utilizan las ecuaciones de movimiento para un fluido incompresible considerando que la velocidad del fluido es $\mathbf{u} = (u_1, u_2, u_3)$, que la masa del fluido es fija en el volumen, definiendo una presión p y una constante de densidad ρ. Por lo tanto el volumen del fluido V puede cambiar solamente debido a la variación del flujo del fluido en una sección transversal en un tiempo determinado, teniendo entonces una igualdad en ambos volúmenes la cual dependerá de las variables mencionadas, siendo lo anterior representado mediante la siguiente expresión.

$$\frac{d}{dt}\int_V \rho dV = -\int_S \rho(\mathbf{u}\cdot\mathbf{n})dS = 0 \tag{3.1}$$

Donde S es la superficie del volumen V y $\mathbf{n} = (n_1, n_2, n_3)$ es la unidad normal saliente de volumen. Similarmente el momento del fluido en un dominio fijo V puede cambiar únicamente en respuesta a las fuerzas aplicadas al flujo que cruza los límites del dominio. Entonces para un fluido viscoso está implícita la conservación del momento, de tal forma que se cumple la condición siguiente.

$$\frac{d}{dt}\int_V \rho u_i dV = -\int_S [(\mathbf{u}\cdot\mathbf{n})\rho u_i + p n_i]dS = 0 \tag{3.2}$$

Usando el teorema de la divergencia se pueden convertir las integrales de superficie en integrales de volumen obteniendo las siguientes expresiones.

$$\int_V (\rho\frac{\partial u_i}{\partial t} + \rho\nabla\cdot(u_i\mathbf{u}) + \frac{\partial p}{\partial x_i})dV = 0 \tag{3.3}$$

$$\int_V \nabla\cdot\mathbf{u}dV = 0 \tag{3.4}$$

Al considerar que el volumen es arbitrario se presentan las dos condiciones siguientes.

$$\rho\frac{\partial\mathbf{u}}{\partial t} + \rho(\nabla\cdot\mathbf{u})\mathbf{u} + \nabla p = 0 \tag{3.5}$$

$$\nabla \cdot \mathbf{u} = 0 \tag{3.6}$$

Cuando los movimientos del fluido se presentan en amplitudes pequeñas, se pueden ignorar los términos no lineales en la cóclea, debido a ésto se modifican las condiciones anteriores dando como resultado lo siguiente.

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \nabla p = 0 \tag{3.7}$$

$$\nabla \cdot \mathbf{u} = 0 \tag{3.8}$$

Un caso especial muy importante es cuando $\mathbf{u} = \nabla\phi$ para cualquier potencial ϕ en un flujo irrotacional quedando las ecuaciones anteriores de la forma.

$$\rho \frac{\partial \phi}{\partial t} + p = 0 \tag{3.9}$$

$$\nabla^2 \phi = 0 \tag{3.10}$$

3.2. La membrana basilar como un sistema de osciladores armónicos forzados

En su trabajo Lesser y Berkeley [Les72] consideran que una forma de modelar el comportamiento físico de la cóclea es combinar las dos ecuaciones anteriores con la ecuación de un oscilador armónico forzado. Si se considera el valor de la presión como una constante arbitraria en los compartimientos superior e inferior de la membrana basilar, se tienen dos copias de las dos ecuaciones dadas en la forma siguiente.

$$\rho \frac{\partial \phi_1}{\partial t} + p_1 = \rho \frac{\partial \phi_2}{\partial t} + p_2 = 0 \tag{3.11}$$

$$\nabla^2 \phi_1 = \nabla^2 \phi_1 = 0 \tag{3.12}$$

Pudiendo entonces cada punto de la membrana basilar ser modelado como un oscilador armónico forzado con masa, resistencia mecánica y constante de elasticidad, los cuales varian a lo largo de la membrana. Debido a esta condición, el movimiento de cada una de las partes de la membrana se considera que es independiente del movimiento de sus partes vecinas, suponiendo que no hay acoplamiento lateral [Les72]. La deflexión de la membrana se puede considerar entonces como una ecuación de onda unidimensional $\eta(x,t)$ la cual es considerada como la solución para la ecuación del oscilador armónico forzado [Elm85], siendo su expresión matemática la siguiente.

$$m(x)\frac{\partial^2 \eta}{\partial t^2} + R_m(x)\frac{\partial \eta}{\partial t} + k(x)\eta = p_2(x, \eta(x,t), t) - p_1(x, \eta(x,t), t) \tag{3.13}$$

Donde $m(x)$ es la masa por unidad de área de la membrana basilar, $R_m(x)$ es la resistencia mecánica y $k(x)$ es la constante de elasticidad. Debido a que el desplazamiento de la membrana es pequeño, la fuerza que se considera se toma como la diferencia de presión entre el punto $y = 0$ y el punto máximo $y = \eta$. Lo cual simplifica en forma significativa a la ecuación del oscilador armónico pudiendo especificar las condiciones límites del modelo.

$$m(x)\frac{\partial^2\eta}{\partial t^2} + R_m(x)\frac{\partial\eta}{\partial t} + k(x)\eta = p_2(x,0,t) - p_1(x,0,t) \tag{3.14}$$

Si $\partial\phi/\partial y$ es la componente en y de la velocidad del fluido, las condiciones límites de la membrana basilar se pueden considerar dadas de la forma.

$$\frac{\partial\eta}{\partial t} = \frac{\partial\phi_1}{\partial y} = \frac{\partial\phi_2}{\partial y} \qquad y = 0, \quad 0 \leq x \leq L \tag{3.15}$$

Se puede considerar además que no hay movimiento vertical en la parte superior, teniendo entonces la condición siguiente.

$$\frac{\partial\phi_1}{\partial y} = 0 \qquad y = l, \quad 0 \leq x \leq L \tag{3.16}$$

Aunque existen muchas formas de poder excitar externamente un sistema de un oscilador forzado, se considera que el sistema es excitado por un movimiento del estribo en contacto con la ventana oval. Siendo $\partial\phi/\partial x$ la componente de x de la velocidad del fluido, teniendo entonces cuando $x = 0$ la siguiente condición límite.

$$\frac{\partial\phi_1}{\partial x} = \frac{\partial F(y,t)}{\partial t} \qquad 0 \leq y \leq l \tag{3.17}$$

Donde $F(y,t)$ es el desplazamiento horizontal de la ventana oval, considerando además que no hay movimiento horizontal cerca del final, siendo en este punto $x = l$ y por lo tanto teniendo la siguiente condición.

$$\frac{\partial\phi_1}{\partial x} = 0 \qquad 0 \leq y \leq l \tag{3.18}$$

3.3. Propuesta de solución usando análisis por resonancia

La solución al modelo de la membrana basilar como un sistema de osciladores armónicos forzados, ha sido propuesta en forma numérica a partir del modelado de flujo de potencial por series de Fourier por Lesser y Berkeley en 1972 [Les72]. Posteriormente en 1974 Siebert generaliza la solución de Lesser y Berkeley considerando una fuerza mecánica en los dos extremos de la membrana basilar [Sie74], una solución similar fué encontrada en 1981 por Peskin [Pes81]. Sin embargo los estudios posteriores consideraron la forma de la membrana basilar para dar solución al modelo, destacando los estudios en 1984 de Rhode

[Rho84], en 1985 de Hudspeth [Hud85] y en 1996 de Boer [Boe96]. Recientemente en 2007 Elliott, Ku y Lineton propusieron una solución considerando el modelo de elementos activos de micromecanismos en la biomecánica de la cóclea propuesto anteriormente por Neely [Nee81] haciendo un análisis de espacio estado [Ell07], posteriormente un año después en 2008 ampliaron su solución haciendo un análisis estadístico de las inestabilidades del modelo [Ku08]. Las diferentes soluciones al modelo de Lesser y Berkeley desarrolladas hasta la actualidad presentan diferentes métodos para obtener el comportamiento de la membrana basilar, pero no proporcionan una relación directa entre sus características físicas y la distancia a lo largo de la membrana donde se presenta la máxima amplitud para cada frecuencia de excitación.

En este trabajo de Tesis se propone el empleo de una solución alternativa de la ecuación del oscilador forzado descrita por la ecuación 3.13 utilizando el análisis por resonancia. Su aportación respecto a las diferentes soluciones ya encontradas es poder determinar un único valor de amplitud máxima de resonancia en cada sección a lo largo de la membrana basilar para cada frecuencia de excitación, obteniendo una relación frecuencia distancia la cual sólo depende de las características físicas de resistencia mecánica, constante de elasticidad y masa por unidad de área a lo largo de la membrana, lo cual coincide con la teoría de los puntos de audición de Békésy [Bek60].

Para realizar el análisis se considera a cada sección de la membrana como un oscilador armónico forzado aislado, el cual es excitado por una fuerza externa $Fe^{j\omega t}$ que representa la fuerza excitadora sobre cada sección de la membrana basilar, esta fuerza es producida por las vibraciones transmitidas al interior de la cóclea por la ventana oval, la ecuación diferencial que describe el movimiento resultante del sistema queda de la siguiente forma.

$$m(x)\frac{d\eta}{dt^2} + R_m(x)\frac{d\eta}{dt} + k(x)\eta = Fe^{j\omega t} \tag{3.19}$$

Donde F es la magnitud de la fuerza excitadora y ω es su frecuencia angular, lo cual implica que la fuerza excitadora $Fe^{j\omega t}$ es periódica y de forma compleja [Den85], por lo tanto se puede considerar que el desplazamiento η también es complejo, estando entonces la solución de la ecuación diferencial definida por el desplazamiento [Wyl75]. Siendo el desplazamiento complejo $\eta = \mathbf{A}e^{j\omega t}$ donde $\mathbf{A}$ es la amplitud compleja, al sustituir el desplazamiento, su primera derivada y su segunda derivada, la ecuación anterior se expresa de la siguiente forma.

$$m(x)(j^2\omega^2\mathbf{A}e^{j\omega t}) + R_m(x)(j\omega\mathbf{A}e^{j\omega t}) + k(x)(\mathbf{A}e^{j\omega t}) = Fe^{j\omega t} \tag{3.20}$$

Si se considera como factor común al desplazamiento se obtiene la siguiente expresión general.

$$(-m(x)\omega^2 + jR_m(x)\omega + k(x))\mathbf{A}e^{j\omega t} = Fe^{j\omega t} \tag{3.21}$$

A partir de la ecuación anterior se obtiene una expresión que modela al desplazamiento en función de los otros términos.

$$\eta = \frac{1}{j\omega} \frac{Fe^{j\omega t}}{R_m(x) + j(\omega m(x) - \frac{k(x)}{\omega})} \tag{3.22}$$

La ecuación anterior puede ser expresada en una forma más simple si se define la impedancia mecánica compleja de entrada del sistema $\mathbf{Z_m}(x)$ como la suma de una parte real dada por la resistencia mecánica y una parte imaginaria dada por la reactancia mecánica de la forma siguiente [Alo67].

$$\mathbf{Z_m}(x) = R_m(x) + jX_m(x) \tag{3.23}$$

Donde la reactancia mecánica $X_m(x)$ se define de la forma.

$$X_m(x) = \omega m(x) - \frac{k(x)}{\omega} \tag{3.24}$$

La impedancia mecánica también puede ser expresada en forma polar $\mathbf{Z_m}(x) = Z_m(x)e^{j\Theta(x)}$ teniendo entonces una expresión en términos de los componentes de magnitud y ángulo de fase.

$$Z_m(x) = \sqrt{R_m(x)^2 + X_m(x)^2} \tag{3.25}$$

$$\Theta(x) = tan^{-1} \frac{X_m(x)}{R_m(x)} \tag{3.26}$$

A partir de las dos ecuaciones anteriores se expresa el desplazamiento complejo en la forma siguiente.

$$\eta = \frac{1}{j\omega} \frac{Fe^{j\omega t}}{Z_m(x)e^{j\Theta(x)}} \tag{3.27}$$

La ecuación anterior se simplifica en un sólo termino exponencial reduciéndola algebraicamente.

$$\eta = \frac{1}{j\omega} \frac{F}{Z_m(x)} e^{j(\omega t - \Theta(x))} \tag{3.28}$$

A continuación se procede a obtener su parte real y su parte imaginaria utilizando la identidad de Euler dada por $e^{j\theta} = cos\theta + jsen\theta$ [Boa06], quedando la ecuación siguiente.

$$\eta = \frac{1}{j\omega} \frac{F}{Z_m(x)} [cos(\omega t - \Theta(x)) + jsen(\omega t - \Theta(x))] \tag{3.29}$$

Estando el desplazamiento de cada sección de la membrana definido por la parte real de la ecuación.

$$\eta = \frac{1}{\omega} \frac{F}{Z_m(x)} sen(\omega t - \Theta(x)) \tag{3.30}$$

La amplitud en la ecuación anterior está definida por $A = \frac{F}{\omega Z_{m(x)}}$ y puede ser expresada algebraicamente en términos de la masa, la resistencia mecánica y la constante de elasticidad en la forma siguiente.

$$A = \frac{F}{\omega\sqrt{R_m(x)^2 + (\omega m(x) - \frac{k(x)}{\omega})^2}} \tag{3.31}$$

La ecuación anterior se puede reacomodar dividiendo ambos factores por la masa y factorizando el denominador, quedando la siguiente ecuación.

$$A = \frac{F/m(x)}{\sqrt{(\omega^2 - \frac{k(x)}{m(x)})^2 + \omega^2 \frac{R_m(x)^2}{m(x)^2}}} \tag{3.32}$$

La ecuación resultante muestra que la amplitud para cada sección de la membrana basilar depende de la frecuencia angular ω de la fuerza aplicada. La amplitud tiene un máximo cuando el denominador tiene su valor mínimo, esto sucede a una frecuencia específica de excitación llamada frecuencia angular de resonancia ω_R, la cual está definida por los valores a lo largo de la membrana basilar de la masa, la constante de elasticidad y la resistencia mécanica [Set71], estando determinada por la ecuación.

$$\omega_R = \sqrt{\frac{k(x)}{m(x)} - \frac{R_m(x)^2}{2m(x)^2}} \tag{3.33}$$

Cuando la frecuencia angular de excitación ω es igual a la frecuencia angular de resonancia ω_R se dice que hay resonancia en amplitud [Alo67]. Cuanto menor es el amortiguamiento más pronunciada es la resonancia y cuando la resistencia mecánica es cero la amplitud de resonancia es infinita, lo cual está determinado por la condición $\omega = \omega_R = \sqrt{k(x)/m(x)}$. La ecuación 3.32 también puede ser expresada como una función de la amplitud que depende de la frecuencia y la distancia para lo cual se considera que $\omega = 2\pi f$, quedando de la siguiente forma.

$$A(x, f) = \frac{F/m(x)}{\sqrt{(4\pi^2 f^2 - \frac{k(x)}{m(x)})^2 + 4\pi^2 f^2 \frac{R_m(x)^2}{m(x)^2}}} \tag{3.34}$$

3.4. Membragrama usando análisis por resonancia

La ecuación 3.33 representa en forma gráfica la distancia sobre la membrana basilar correspondiente a cada frecuencia de resonancia, lo cual se define como un *membragrama*. Para el desarrollo de esta herramienta se usan los parámetros de Neely [Nee81] por ser los más próximos a la respuesta física de la membrana basilar. Al sustituir en la ecuación 3.33 los valores correspondientes de masa, resistencia mecánica, la constante de elasticidad modelada por la ecuación 2.7 y considerando que $\omega_R = 2\pi f_R$ se obtiene la expresión.

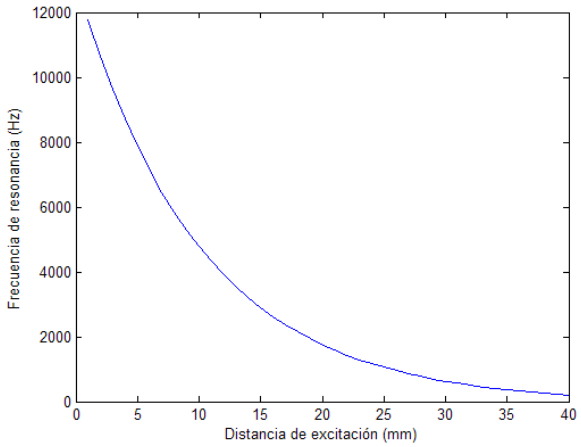


Figura 3.2: Membragrama distancia vs. frecuencia.

$$f_R = \frac{1}{2\pi}\sqrt{\frac{1 \cdot 10^9 e^{-2x}}{0{,}15} - \frac{200^2}{2(0{,}15)^2}} \tag{3.35}$$

Siendo esta ecuación una función entre la distancia y la frecuencia de resonancia, en la figura 3.2 se muestra el membragrama para los límites de distancia de 0.1 mm a 40 mm.

Para obtener una relación de la distancia en función de la frecuencia se despeja a la variable x de la ecuación 3.33 quedando la expresión.

$$x_R = \frac{ln[\frac{(2\pi f_R)^2 m(x) + \frac{R_m(x)^2}{2m(x)}}{10^9}]}{-2} \tag{3.36}$$

Si se sustituyen los parámetros de Neely en la ecuación resultante se obtiene la función entre la frecuencia de resonancia y la distancia.

$$x_R = \frac{ln[\frac{(2\pi f_R)^2 (0{,}15) + \frac{200^2}{2(0{,}15)}}{10^9}]}{-2} \tag{3.37}$$

A partir de esta ecuación se grafica el comportamiento de la membrana basilar de frecuencia de resonancia contra distancia, en la figura 3.3 se muestra el resultado obtenido para el intervalo de frecuencias de 1 Hz a 12000 Hz. La aplicación de esta herramienta al análisis de voz permite determinar las zonas de mayor actividad a lo largo de la membrana basilar, si se tiene una representación en el dominio de la frecuencia de una señal de voz $X(\omega)$ y se considera un límite de energía de amplitud en la transformada de Fourier, se delimitan los componentes más significativos pudiendo hacer su representación en el dominio

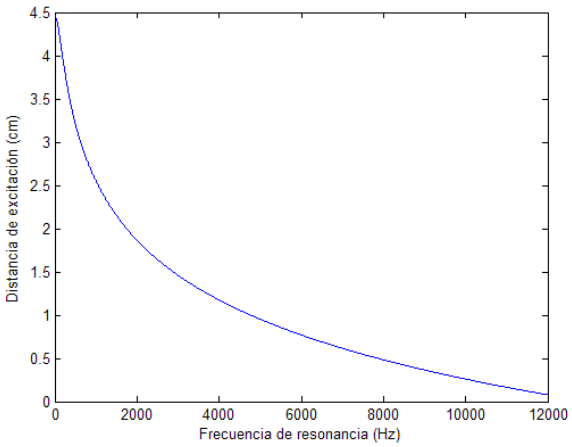


Figura 3.3: Membragrama frecuencia vs. distancia.

de la distancia $X(d)$ mediante el membragrama, lo cual es concordante con la teoría de los puntos de audición de Békésy.

3.5. Parametrización de la voz usando análisis por resonancia

La parametrización de la señal de voz en un vector reducido de coeficientes es necesaria debido a que las representaciones en el dominio del tiempo y de la frecuencia consideran ventanas de análisis muy grandes que resultan ser inadecuadas para su utilización en procesos de reconocimiento. En este trabajo de Tesis se hace la propuesta de utilizar el análisis por resonancia usando los parámetros físicos de la membrana basilar propuestos por Neely junto con el método de diferencias finitas, para el diseño de un arreglo de filtros similares a los usados en los MFCC que representen en el dominio de la frecuencia el comportamiento mecánico acústico de la cóclea y fisiológicamente modelen en forma discreta la actividad de la membrana basilar en distancias equidistantes.

Dependiendo del contenido en frecuencias de una señal de entrada se puede acotar el intervalo en frecuencia de los filtros, estando limitados por una frecuencia máxima *fmax* cercana a la ventana oval y una frecuencia mínima *fmin* cercana al helicotrema. A continuación se obtiene la distancia de resonancia para estas frecuencias límites usando la ecuación 3.37, obteniendose un valor de distancia mínimo *dmin* para la frecuencia máxima y un valor de distancia máximo *dmax* para la frecuencia mínima. Posteriormente la longitud obtenida entre las dos distancias límites se divide en un número de intervalos igualmente espaciados *nint* que corresponde al número total de filtros, para lo cual se utiliza

Tabla 3.1: Filtros cocleares (1 Hz - 4600 Hz)

Frecuencia Hz	Distancia cm	Frecuencia Hz	Distancia cm
1	4.4613	975	2.5785
96	4.2901	1161	2.4073
149	4.1190	1381	2.2361
201	3.9478	1641	2.0650
257	3.7766	1950	1.8938
320	3.6055	2316	1.7226
391	3.4343	2750	1.5515
474	3.2631	3265	1.3803
571	3.0920	3875	1.2091
684	2.9208	4600	1.0380
817	2.7496		

el método de diferencias finitas [Sch88] [Cha02]. Estando lo anterior expresado de la forma.

$$d(n) = dmax + \sum_{n=0}^{n=nint} n\frac{dmin - dmax}{nint + 1}, \tag{3.38}$$

Donde los valores de $d(n)$ son las distancias límites de los intervalos. Una vez conocidos estos valores se realiza su transformación del dominio de la distancia $X(d)$ al domino de la frecuencia $X(\omega)$ mediante la ecuación 3.35, el conjunto de valores obtenidos se utilizan como frecuencias centrales para el arreglo de filtros. En la tabla 3.1 se muestran las frecuencias y las distancias obtenidas para el diseño de un arreglo de 20 filtros en un intervalo de frecuencias de 1 Hz a 4600 Hz y en la figura 3.4 se muestra la gráfica del resultado obtenido. La metodología restante para la obtención del vector paramétrico es similar a la descrita en la sección 2.4 para los MFCC.

A continuación en las figuras 3.5, 3.6 y 3.7 se muestra el proceso de filtrado coclear para los fonemas /s/, la primera vocal /a/ y /l/ en la palabra *sala*, la cual es muestreada a una frecuencia de 16 kHz y digitalizada a 16 bits en formato monofónico. En ambas figuras se observan primero las muestras de la señal de voz capturada, a continuación su transformada de Fourier y por último el resultado del filtrado coclear con un total de 20 filtros diseñados en un intervalo de frecuencias entre 1 Hz y 8 kHz.

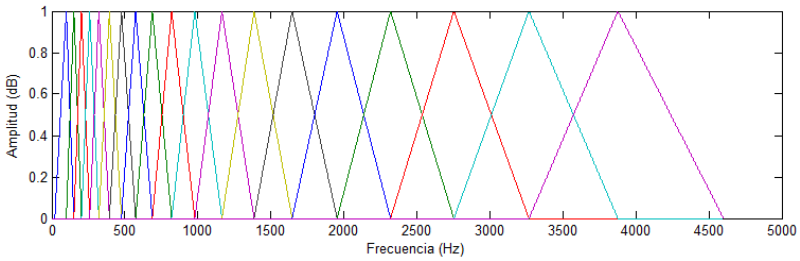


Figura 3.4: Filtros cocleares (1 Hz - 4600 Hz).

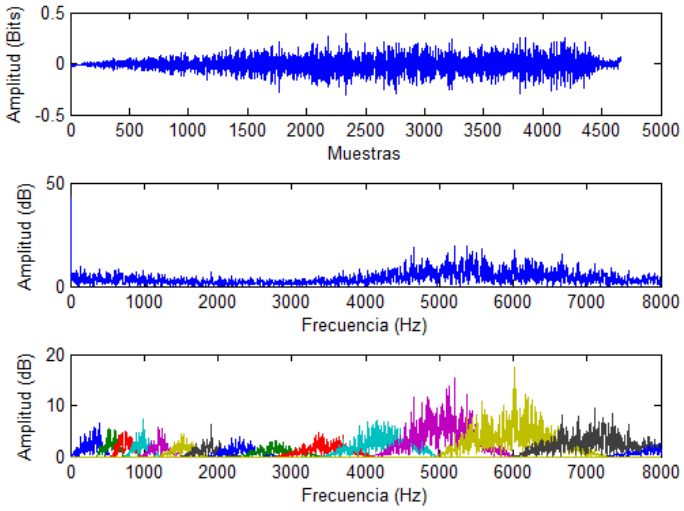


Figura 3.5: Filtrado coclear del fonema /s/ en la palabra *sala*.

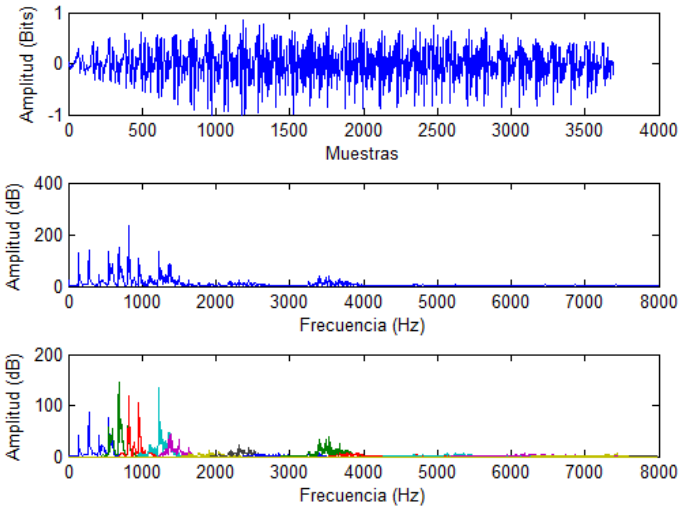


Figura 3.6: Filtrado coclear del fonema /a/ en la palabra *sala*.

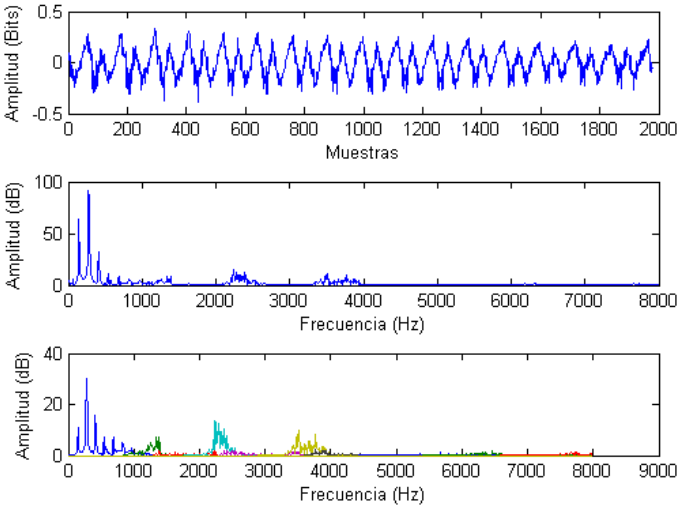


Figura 3.7: Filtrado coclear del fonema /l/ en la palabra *sala*.

Capítulo 4

Experimentos y resultados

Simpre desconfié de los juicios prematuros, porque inconscientemente nos obligan a ajustar los hechos a la teoría y no la teoría a los hechos...
Artur Conan Doyle. Aventuras de Sherlock Holmes, Londres 1891.

En el presente capítulo se muestran los experimentos realizados y los resultados obtenidos. Primero se evalúa el modelo de análisis por resonancia comparándolo con los modelos de Peterson y Bogert, Lesser y Berkeley, Allen, Neely y los resultados experimentales de Békésy. A continuación se utiliza el membragrama para determinar las zonas de mayor actividad a lo largo de la membrana basilar para las señales de voz de los fonemas del español. Por último se muestra la obtención de un vector paramétrico de una señal de voz usando el arreglo de filtros cocleares desarrollado y su implementación en procesos de reconocimiento de voz usando HTK.

4.1. Evaluación del modelo de análisis por resonancia

El modelo de la membrana basilar usando ańalisis por resonancia se compara con los resultados obtenidos en los trabajos de Peterson y Bogert [Pet50], Lesser y Berkeley [Les72], Allen [All77], Neely [Nee81] y los resultados experimentales de Békésy [Bek60]. Se utilizan en cada experimento los parámetros de la membrana basilar de masa m, resistencia mecánica R_m y constante de elasticidad k propuestos por cada investigador y las mismas frecuencias de evaluación. En todos los experimentos el valor de la magnitud de la fuerza de excitación externa F en la ecuación de resonancia 3.34 se considera normalizado, debido a que su variación no modifica la posición a lo largo de la membrana basilar donde se obtiene el máximo valor de amplitud de resonancia.

Primero se hace la comparación entre el análisis por resonancia y el modelo

Tabla 4.1: Análisis por resonancia vs. Peterson y Bogert.

Frecuencia de evaluación (Hz)	Peterson Distancia (cm)	Resonancia Distancia (cm)
31.6	3.125	6.340*
100	3.250	5.159*
316	2.416	4.011*
1000	2.833	2.860
3160	1.700	1.709
10000	0.562	0.557

de Peterson y Bogert, debido a que este modelo no presenta resistencia mecánica ($R_m = 0$) la ecuación 3.34 queda de la forma.

$$A = \frac{F/m(x)}{\sqrt{(\omega^2 - \frac{k(x)}{m(x)})^2}} \tag{4.1}$$

En la ecuación obtenida 4.1 se sustituyen los valores de los párametros de Peterson y Bogert, teniendo la masa un valor de $m = 0{,}143\ g/cm^2$ y estando la constante de elasticidad dada por la ecuación 2.1, obteniendo la función entre la distancia x a lo largo de la membrana basilar donde se presenta la máxima amplitud de resonancia y la frecuencia de excitación externa f de la forma siguiente.

$$A(x, f) = \frac{1/0{,}143}{\sqrt{((2\pi f)^2 - \frac{1{,}72 \cdot 10^9 e^{-2x}}{0{,}143})^2}} \tag{4.2}$$

La ecuación resultante 4.2 se evalúa en el intervalo de frecuencias de 20 Hz a 12000 Hz en incrementos de 1 Hz y en el intervalo de distancias de 0.001 cm a 3.300 cm en incrementos de 0.001 cm, obteniendo para cada distancia un valor de amplitud máxima el cual depende de la frecuencia de excitación, en la tabla 4.1 se muestran los resultados para las frecuencias de evaluación del modelo de Peterson y Bogert y se comparan con los resultados del análisis por resonancia.

En la figura 4.1 se presenta la gráfica obtenida para la frecuencia de 1000 Hz, como se observa la respuesta de resonancia carente de resistencia mecánica es cerrada en forma abrupta en el eje de la frecuencia. Con el objetivo de una visualización apropiada se presentarán en la figura 4.2 las tres frecuencias de evaluación más bajas y en la figura 4.3 las tres frecuencias de evaluación más altas. La comparación entre los resultados proporcionados por el análisis de resonancia resultan exitosos teniendo total concordancia con el método de integración numérica propuesto por Peterson y Bogert en su intervalo de trabajo útil de frecuencias altas. En la parte de las frecuencias bajas donde el método de Peterson y Bogert no proporciona resultados satisfactorios el análisis por

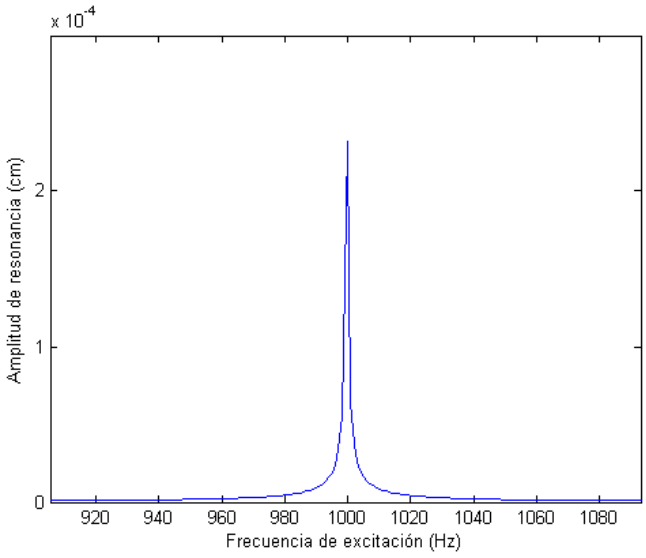


Figura 4.1: Análisis por resonancia para 1000 Hz (Parámetros de Peterson).

resonancia puede ser extendido a estas frecuencias bajas aunque los resultados no coincidan con los valores de las medidas físicas reportadas de la cóclea*.

A continuación el análisis por resonancia se compara con el modelo desarrollado por Lesser y Berkeley, en la ecuación 3.30 se sustituye el parámetro de la masa de $m = 0{,}05\ g/cm^2$, la ecuación de la constante de elasticidad 2.3 y la ecuación de la resistencia mecánica 2.4 descritas en su trabajo, quedando una ecuación que es función entre la distancia x a lo largo de la membrana basilar donde se presenta la máxima amplitud de resonancia y la frecuencia de excitación f de la forma.

$$A(x,f) = \frac{1/0{,}05}{\sqrt{((2\pi f)^2 - \frac{10^9 e^{-3x}}{0{,}05})^2 + (2\pi f)^2 \frac{(3000e^{-1{,}5x})^2}{0{,}05^2}}} \tag{4.3}$$

La ecuación resultante 4.3 se evalúa en el intervalo de distancias de 0.001 cm a 3.600 cm en incrementos de 0.001 cm y el intervalo de frecuencias de excitación de 20 Hz a 1000 Hz en intervalos de 1 Hz. En la tabla 4.2 se puede observar la comparación entre los resultados de ambos modelos para las mismas frecuencias de evaluación.

Debido a que el modelo de Lesser y Berkeley es una aproximación por series de Fourier de la respuesta de la onda envolvente que se propaga sobre la membrana basilar, los resultados obtenidos respecto al análisis por resonancia son diferentes. La figura 4.4 muestra la gráfica obtenida para la frecuencia de 800 Hz, se observa claramente la diferencia entre un modelo con resistencia

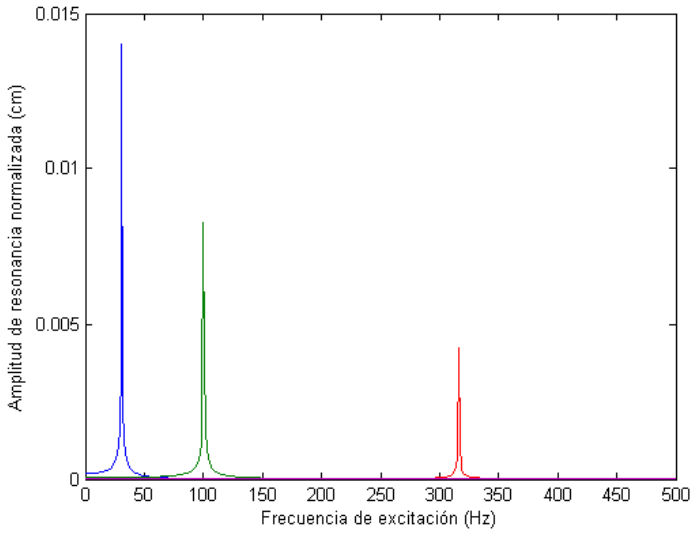


Figura 4.2: Análisis por resonancia para 31.6 Hz, 100 Hz y 316 Hz (Parámetros de Peterson).

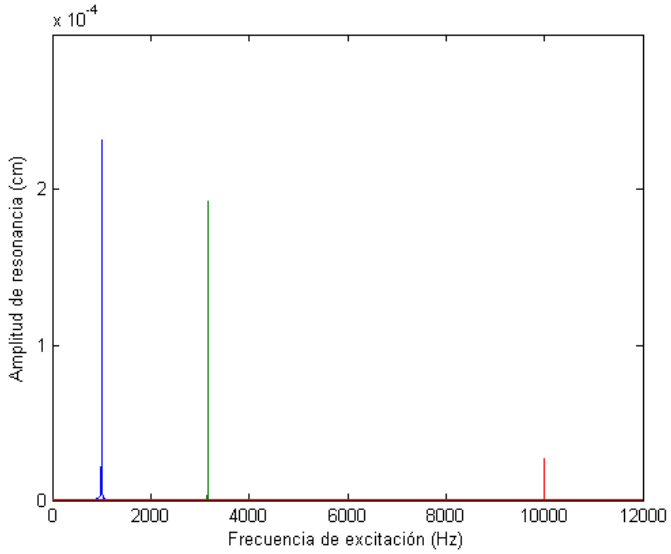


Figura 4.3: Análisis por resonancia para 1000 Hz, 3160 Hz y 10000 Hz (Parámetros de Peterson).

Tabla 4.2: Análisis por resonancia vs. Lesser y Berkeley.

Frecuencia de evaluación (Hz)	Lesser Distancia (cm)	Resonancia Distancia (cm)
100	3.41	3.580
200	2.88	3.116
400	2.26	2.655
800	1.72	2.193

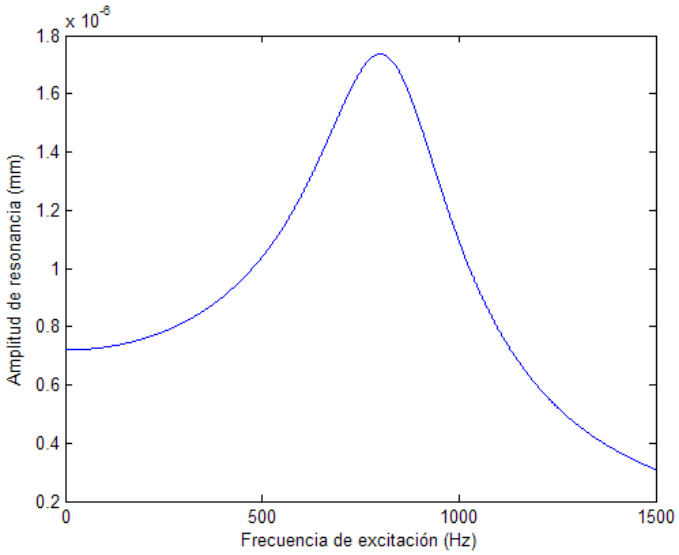


Figura 4.4: Análisis por resonancia para 800 Hz (Parámetros de Lesser).

mecánica como el propuesto por Lesser y Berkeley y un modelo sin resistencia mecánica como el de Peterson y Bogert. En la figura 4.5 se muestran las gráficas de las cuatro frecuencias de evaluación con las cuales se comparan ambas metodologías, se puede observar que a pesar de la diferencia entre los resultados las curvas de resonancia tienen un comportamiento similar al de la envolvente de la serie de Fourier mostrada en la figura 2.10.

El tecer modelo con el cual se compara la solución del análisis por resonancia es el de Allen, la ecuación 3.34 se evalua con los parámetros descritos en su trabajo considerando el valor de la masa de $m = 0{,}1\ g/cm^2$ y el valor de $a = 1{,}5$ en la ecuación de la constante de elasticidad 2.5 y en la ecuación de la resistencia mecánica 2.6, la ecuación resultante 4.4 queda entonces como una función de dos variables dadas por el desplazamiento a lo largo de la membrana basilar x y la frecuencia de excitación f de la forma.

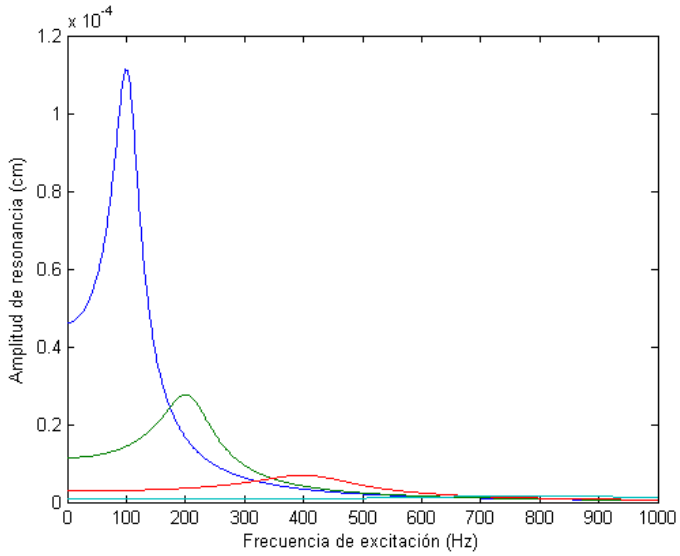


Figura 4.5: Análisis por resonancia para 100 Hz, 200 Hz, 400 Hz y 800 Hz (Parámetros de Lesser).

$$A(x,f) = \frac{1/0{,}1}{\sqrt{((2\pi f)^2 - \frac{10^9 e^{-3x}}{0{,}1})^2 + (2\pi f)^2 \frac{(300e^{-1,5x})^2}{0{,}1^2}}} \tag{4.4}$$

La ecuación se evalúa en el intervalo de frecuencias de excitación de 20 Hz a 12000 Hz en incrementos de 1 Hz y en el intervalo de distancias de 0.0001 cm a 3.5000 cm en incrementos de 0.0001 cm, en la tabla 4.3 se muestran los resultados obtenidos por ambos modelos para las mismas frecuencias.

Se observa igualdad en todos los resultados, lo cual significa que se obtienen los mismos valores utilizando diferentes procesos, teniendo en un caso la función de Green y en el otro el análisis por resonancia. En la figura 4.6 se muestra la gráfica del análisis por resonancia para la frecuencia de 1000 Hz, se observa un comportamiento similar al obtenido al utilizar el análisis por resonancia con los parámetros de Lesser y Berkeley. Con el objetivo de una visualización apropiada en la figura 4.7 se muestran las gráficas de resonancia para las tres frecuencias de evaluación más bajas y en la figura 4.8 se muestran las gráficas de las cuatro frecuencias más altas.

El cuarto modelo con el que se evalúa y compara la solución del análisis por resonancia es el de Neely, la ecuación 3.30 se evalúa utilizando las ecuaciones de sus parámetros de masa de $m = 0{,}15\ g/cm^2$, de resistencia mécanica de $R_m = 200\ dyna \cdot seg/cm^2$ y la ecuación de la constante de elasticidad 2.7, quedando una función que depende de la posición a lo largo de la membrana

Tabla 4.3: Análisis por resonancia vs. Allen.

Frecuencia de evaluación (Hz)	Allen Distancia (cm)	Resonancia Distancia (cm)
100	3.5000	3.61000
200	3.1500	3.15000
500	2.5375	2.53800
1000	2.1000	2.07550
2000	1.5750	1.61350
5000	0.9625	1.00260
10000	0.4666	0.54055

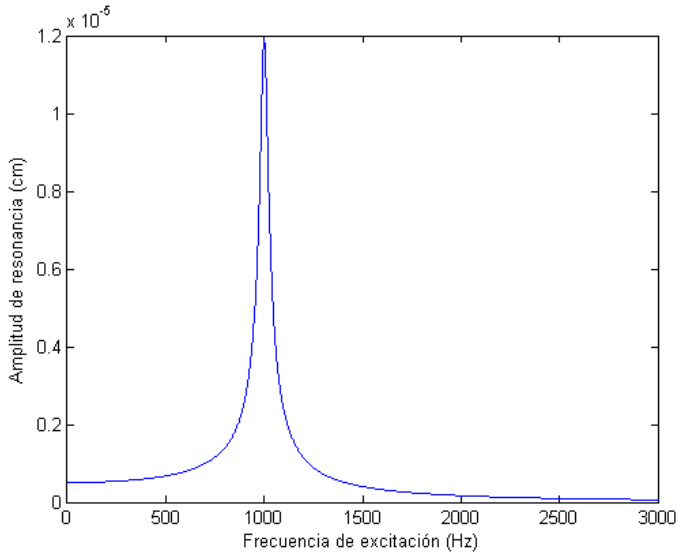


Figura 4.6: Análisis por resonancia para 1000 Hz (Parámetros de Allen).

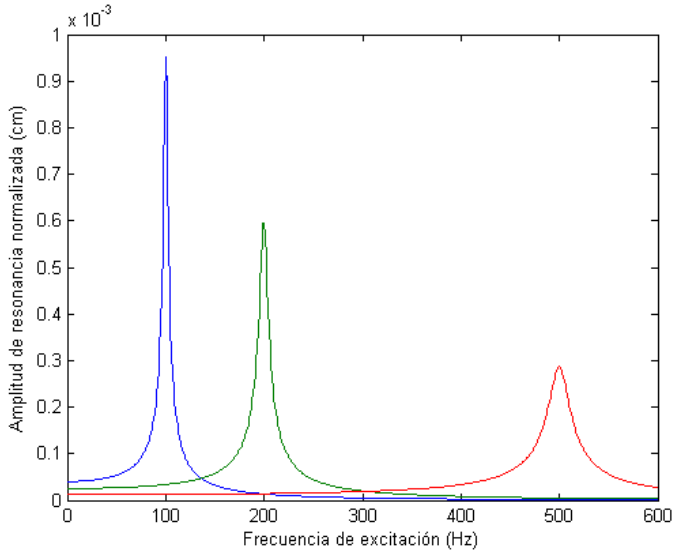


Figura 4.7: Análisis por resonancia para 100 Hz, 200 Hz y 500 Hz (Parámetros de Allen).

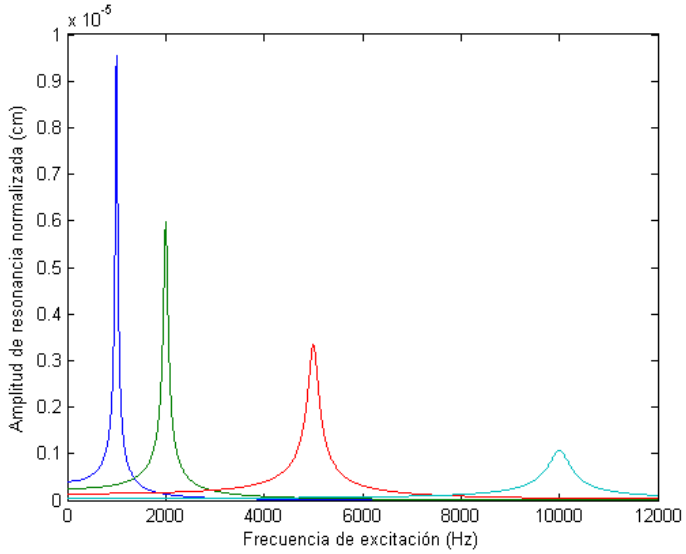


Figura 4.8: Análisis por resonancia para 1000 Hz, 2000 Hz, 5000 Hz y 10000 Hz (Parámetros de Allen).

Tabla 4.4: Análisis por resonancia vs. Neely.

Frecuencia de evaluación (Hz)	Neely Distancia (cm)	Resonancia Distancia (cm)
400	3.250	3.41500
570	2.875	3.09300
800	2.625	2.77000
1130	2.375	2.43400
1600	2.000	2.09000
2260	1.625	1.74700
3200	1.375	1.40020
4520	1.000	1.05560
6390	0.700	0.70950
9040	0.375	0.36281

basilar x y de la frecuencia de excitación f de la forma siguiente.

$$A(x,f) = \frac{1/0{,}15}{\sqrt{((2\pi f)^2 - \frac{10^9 e^{-2x}}{0{,}15})^2 + (2\pi f)^2 \frac{(200)^2}{0{,}15^2}}} \tag{4.5}$$

La ecuación resultante 4.5 se evalúa en el intervalo de frecuencias de 20 Hz a 10000 Hz en incrementos de 1 Hz y en el intervalo de distancias de 0.00001 a 3.50000 en incrementos de 0.00001 cm. En la tabla 4.4 se muestran los resultados de ambos modelos para las mismas frecuencias de evaluación.

Al igual que como sucede con el modelo de Allen existe igualdad entre los resultados obtenidos por ambos procesos, ya sea utilizando el método de diferencias finitas propuesto por Neely o el análisis por resonancia. En la figura 4.9 se muestra la gráfica del análisis por resonancia para la frecuencia de f=1130 Hz y en la figura 4.10 se muestran las gráficas en conjunto para todas las frecuencias de evaluación.

4.2. Análisis de fonemas con el membragrama

En esta sección se presentan los resultados obtenidos usando el membragrama para el análisis de los fonemas del español. Se presentan para cada análisis tres gráficas, primero se muestran las carácteristicas acústicas en el dominio del tiempo, posteriormente su espectro en frecuencia obtenido a partir de la transformada discreta de Fourier y por último el membragrama correspondiente del espectro en frecuencia. Se utiliza un factor de 1/5 de amplitud respecto a la transformada de Fourier para determinar las frecuencias más representativas del espectro y hacer su transformación al dominio de la distancia.

Para la clasificación de los fonemas se tomaron como base los estudios de Antonio Quilis [Qui99] por ser la referencia más completa para el idioma español,

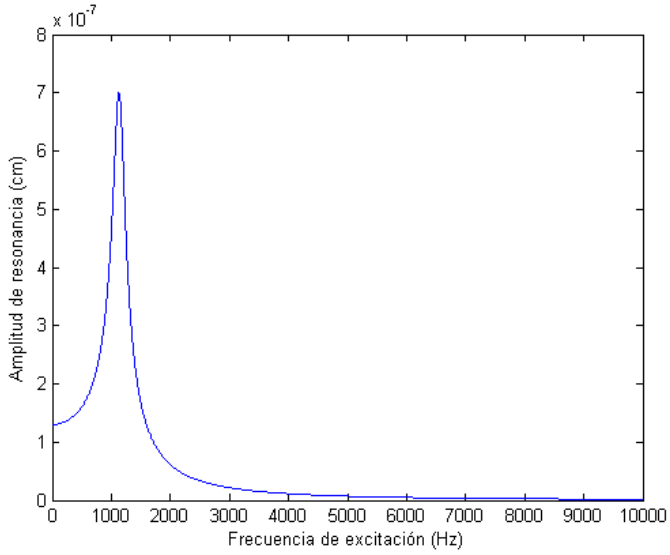


Figura 4.9: Análisis por resonancia para 1130 Hz (Parámetros de Neely).

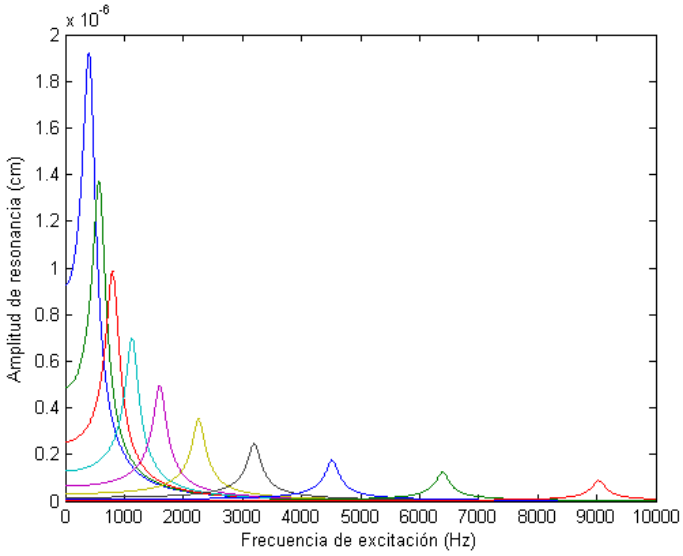


Figura 4.10: Análisis por resonancia para 400 Hz, 570 Hz, 800 Hz, 1130 Hz, 1600 Hz, 2260 Hz, 3200 Hz, 4500 Hz, 6390 Hz y 9040 Hz (Parámetros de Neely).

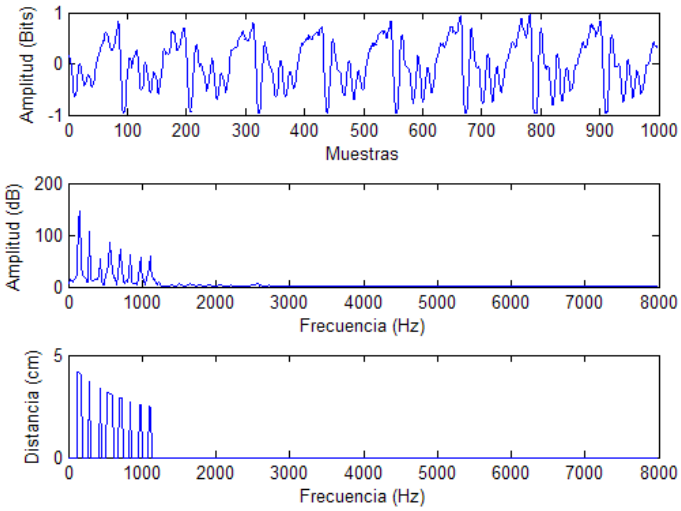


Figura 4.11: Membragrama para fonema vocálico /a/.

las grabaciones de voz para este experimento fueron realizadas por un locutor masculino de 35 años en forma enfatizada con el objetivo de resaltar lo más posible las características acústicas de cada fonema. Los primeros sonidos que se analizan son las cinco vocales las cuales han sido grabadas en forma aislada. Una característica de estos fonemas es la existencia de estructuras de formantes claros debidas al paso del flujo de aire por el conducto bucal sin resistencia y con las cavidades resonadoras potenciando los armónicos distintivos de cada vocal, en las figuras 4.11 a 4.15 se muestran los respectivos análisis.

A continuación se analizan dos casos particulares de las oclusivas sordas o explosivas, siendo este tipo de sonido producido por el cierre (oclusión) de los órganos fonadores durante un intervalo de tiempo, seguido de su apertura con la consiguiente salida brusca de aire (explosión). En la figura 4.16 se presenta el caso de la /p/ con punto de articulación labial al inicio de la palabra *papá* y en la figura 4.17 el caso de la /t/ con punto de articulación dental al inicio de la palabra *tamal*.

Una característica de las consonantes nasales es que en ellas se produce un cierre de los órganos articulatorios bucales con la consiguiente expulsión de aire a través de los conductos nasales, en la figura 4.18 se muestra el resultado obtenido del análisis con el membragrama para la consonante nasal /m/ con punto de articulación labial al inicio de la palabra *mamá*.

Cuando se realiza un estrechamiento entre dos órganos articulatorios en la expiracion de aire se produce la fricación, los fonemas que tienen esta característica se denominan fricativos y se diferencian de los demás por el ruido que

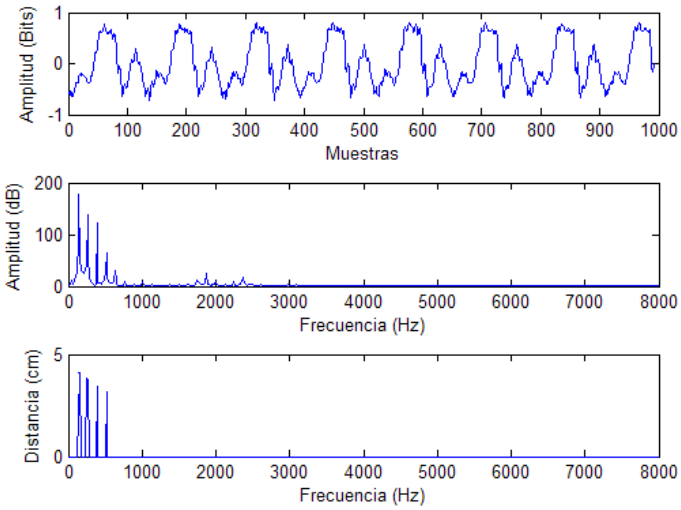


Figura 4.12: Membragrama para fonema vocálico /e/.

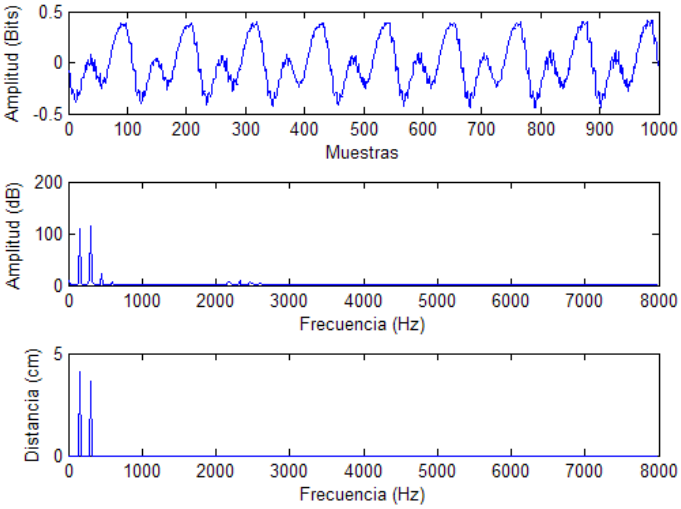


Figura 4.13: Membragrama para fonema vocálico /i/.

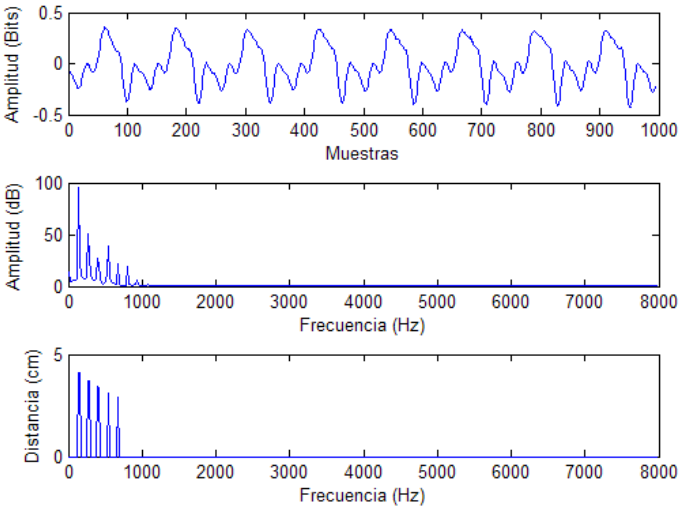


Figura 4.14: Membragrama para fonema vocálico /o/.

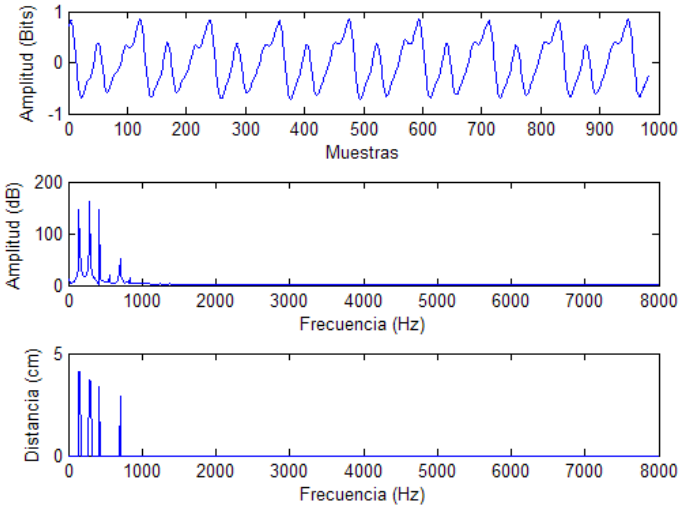


Figura 4.15: Membragrama para fonema vocálico /u/.

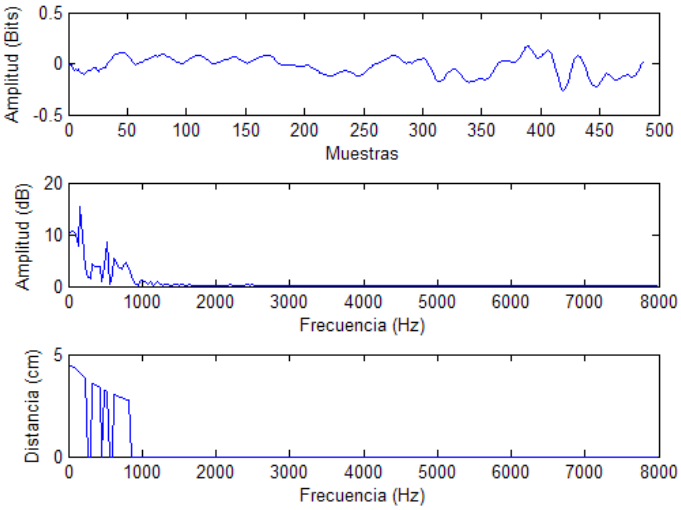


Figura 4.16: Membragrama para fonema oclusivo /p/.

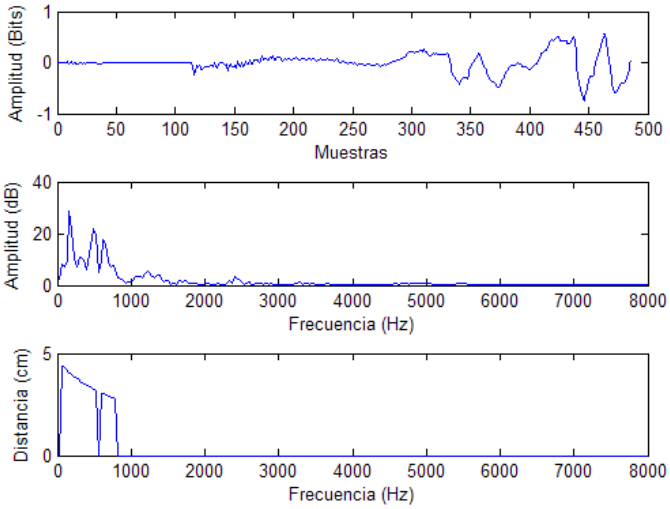


Figura 4.17: Membragrama para fonema oclusivo /t/.

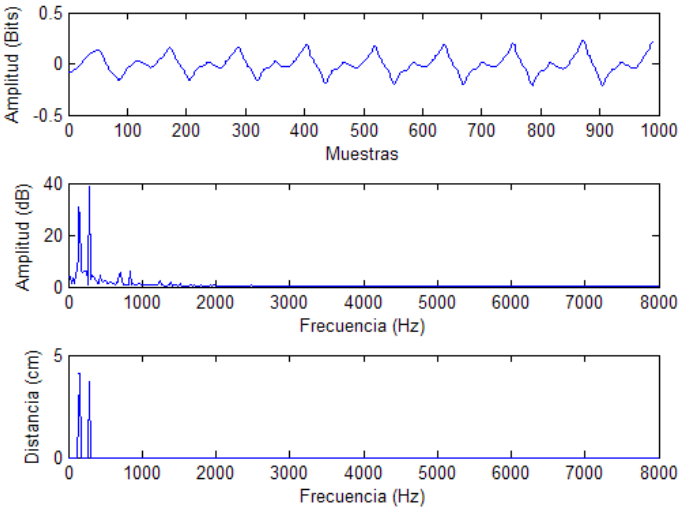


Figura 4.18: Membragrama para fonema nasal /m/.

presentan. Para distingir las fricativas españolas entre si, se recurre a determinar la altura frecuencial a la que se presenta su mayor energía, teniendo la mayor parte de estos sonidos resonancias altas. En la figura 4.19 se presenta el membragrama del fonema /f/ al inicio de la palabra *flecha* el cual se clasifica como fricativo labiodental y sordo debido a que su primera parte corresponde al sonido del aire al paso por una región estrecha, en la figura 4.20 se presenta el membragrama del fonema /s/ al inicio de la palabra *sala* el cual se clasifica como fricativo lingüoalveolar y sordo.

Las consonantes líquidas se producen al pasar el aire por la cavidad bucal con una oclusión central o lateral, de manera que estas consonantes se encuentran acústicamente entre las vocales y las demás consonantes, en la figura 4.21 se muestra el membragrama para el fonema líquido lateral /l/ al inicio de la palabra *limón*. Debido a la poca resistencia a la salida del aire que existe en las consonantes laterales, acústicamente existen formantes similares a los sonidos vocálicos.

Adicionalmente a los fonemas líquidos laterales existen los vibrantes que se producen por medio de interrupciones a la salida del aire, en la figura 4.22 se muestra el membragrama para el fonema líquido vibrante /r̄/ al inicio de la palabra *ratón* en el cual se presentan varias oclusiones seguidas.

Con el objetivo de observar el comportamiento del membragrama en ventanas de análisis grandes se muestra la respuesta del membragrama para los segmentos completos de los fonemas de la palabra *sala*, en la figura 4.23 para la /s/, en la 4.24 para la primera vocal /a/ y en 4.25 para la /l/.

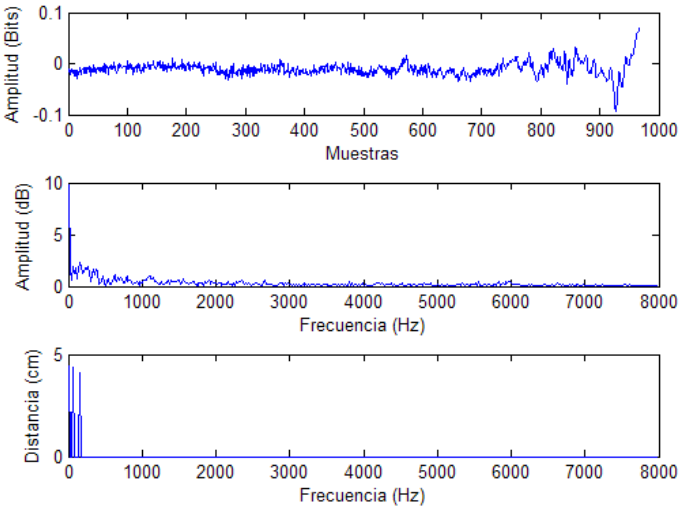


Figura 4.19: Membragrama para fonema fricativo /f/.

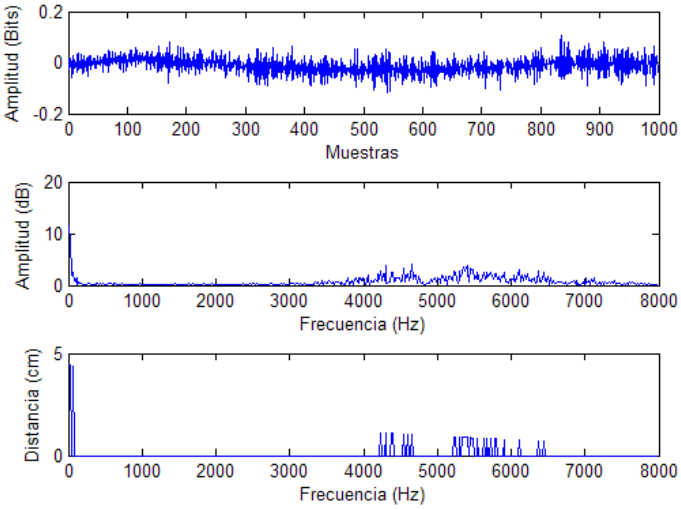


Figura 4.20: Membragrama para fonema fricativo /s/.

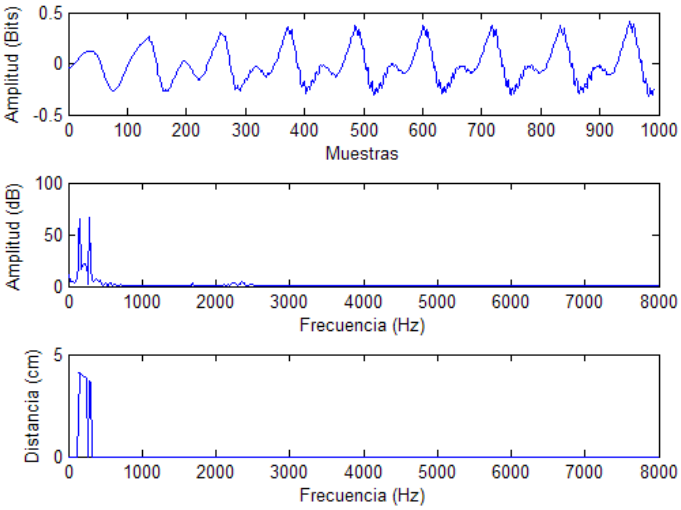


Figura 4.21: Membragrama para fonema semivocálico /l/.

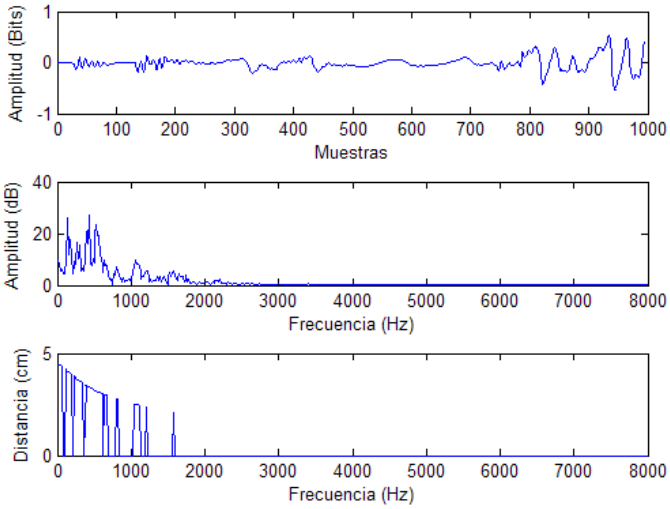


Figura 4.22: Membragrama para fonema semivocálico /r̄/.

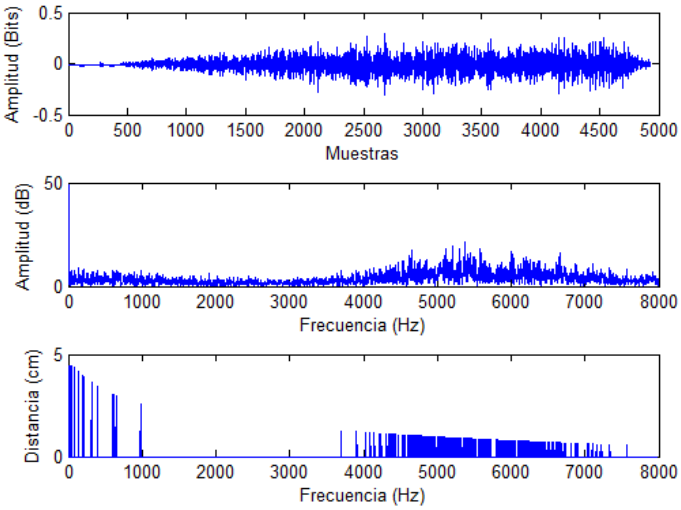


Figura 4.23: Membragrama para fonema /s/ en la palabra *sala*.

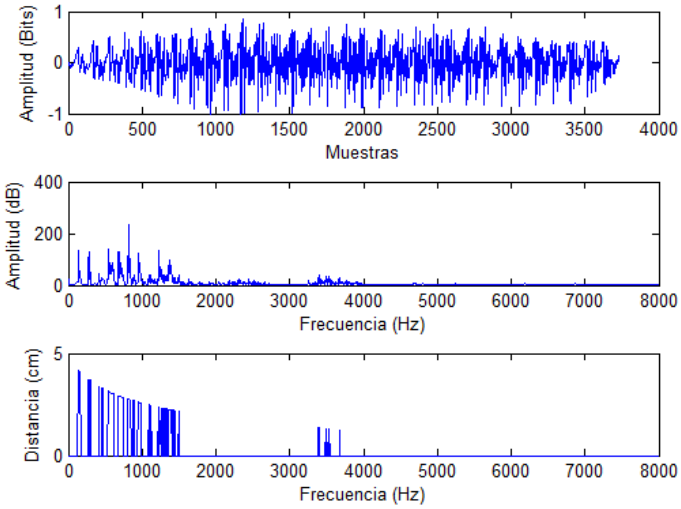


Figura 4.24: Membragrama para el primer fonema /a/ en la palabra *sala*.

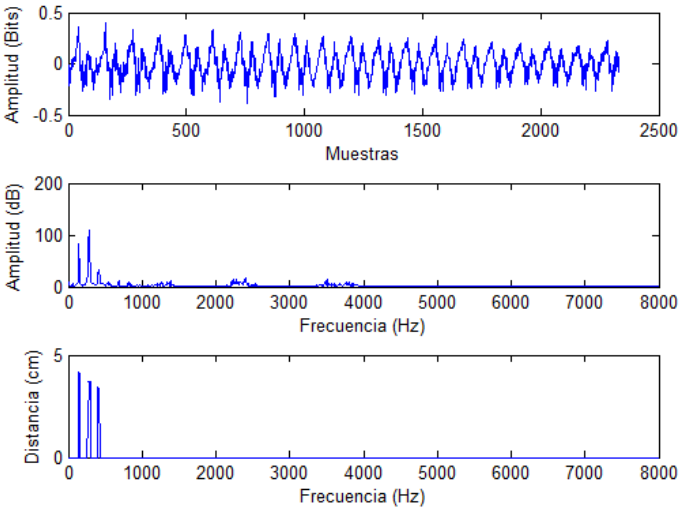


Figura 4.25: Membragrama para fonema /l/ en la palabra *sala*.

4.3. Parámetros cocleares y experimentos de reconocimiento de voz en HTK

El arreglo de filtros cocleares se implementa en HTK modificando únicamente la parte del código donde se genera el arreglo de frecuencias centrales de los filtros con base en la escala de Mel. En el archivo de configuración se establecen los valores de las frecuencias límites superior (HIFREQ) e inferior (LOFREQ) y únicamente se modifica el arreglo de frecuencias centrales del banco de filtros de Mel en el codigo HTK sustituyendo su contenido por los valores de las frecuencias centrales del arreglo de filtros cocleares de la Tabla 3.1 [You09], a continuación en la tabla 4.5 se muestran las características del archivo de configuración para la implementación de los filtros cocleares en los experimentos desarrollados.

Las pruebas desarrolladas se evalúan con un corpus de voz del idioma español de los diez dígitos numéricos pronunciados en forma aislada por un locutor masculino adulto. Las muestras de voz fueron grabadas en un laboratorio de informática con condiciones de ruido controladas enfatizando la propunciación, usando un micrófono de carbón estándar, con una frecuencia de muestreo de 11025 Hz y una cuantificación de 16 bits, teniendo veinte muestras por cada dígito de las cuales se consideran diez para el entrenamiento y diez para el reconocimiento.

Con el objetivo de evaluar la metodología desarrollada se realiza el mismo experimento usando los parámetros LPC y MFCC con un arreglo de veinte

Tabla 4.5: Archivo *configuración* de los filtros cocleares.

Instrucción	Descripción
# coding parameters	Uso de HTK
SOURCEFORMAT = WAV	Formato de audio
TARGETKIND = MFCC	Parámetros MFCC
TARGETRATE = 100000.0	Período de muestreo
SAVECOMPRESSED = T	Uso de HTK
SAVEWITHCRC = T	Uso de HTK
WINDOWSIZE = 300000.0	Tamaño de ventana
USEHAMMING = T	Tipo de ventana
PREEMCOEF = 0.97	Coeficiente de preénfasis
NUMCHANS = 20	Número de filtros
NUMCEPS = 19	Número de coeficientes
LOFREQ = 1	Frecuencia mínima
HIFREQ = 4600	Frecuencia máxima
ENORMALISE = F	Uso de HTK

Tabla 4.6: Pruebas de reconocimiento.

Parámetros	Porcentaje
LPC	99 %
MFCC	100 %
Cocleares	100 %

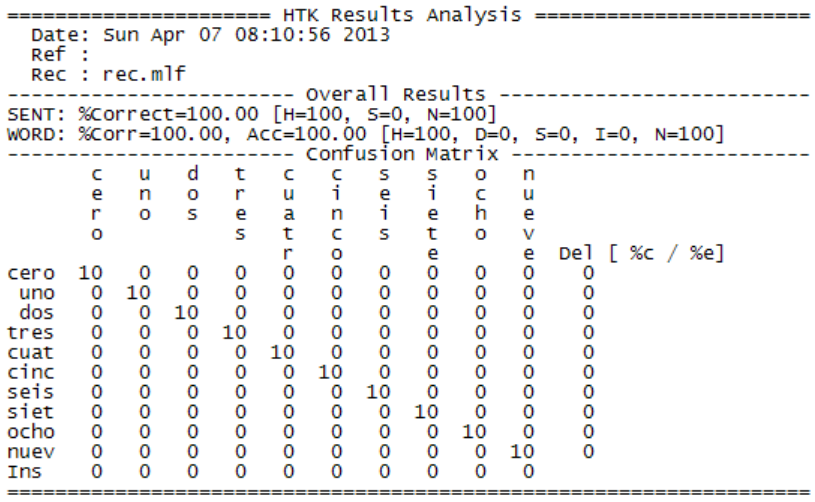
filtros en un intervalo de 1 Hz a 4600 Hz, considerando en los tres experimentos 19 coeficientes de reconocimiento, en las figuras 4.26, 4.27 y 4.28 se muestran los análisis de los resultados proporcionados por HTK para cada una de las metodologías con sus respectivos porcentajes de reconocimiento y en la tabla 4.6 la comparación entre los resultados obtenidos.

```
====================== HTK Results Analysis =======================
  Date: Sat Apr 06 17:05:11 2013
  Ref :
  Rec : rec.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=99.00 [H=99, S=1, N=100]
WORD: %Corr=99.00, Acc=99.00 [H=99, D=0, S=1, I=0, N=100]
------------------------ Confusion Matrix -------------------------
       c   u   d   t   c   c   s   s   o   n
       e   n   o   r   u   i   e   i   c   u
       r   o   s   e   a   n   i   e   h   e
       o           s   t   c   s   t   o   v
                       r   o       e       e   Del [ %c / %e]
cero  10   0   0   0   0   0   0   0   0   0     0
uno    0  10   0   0   0   0   0   0   0   0     0
dos    0   0  10   0   0   0   0   0   0   0     0
tres   0   0   0  10   0   0   0   0   0   0     0
cuat   0   0   0   0  10   0   0   0   0   0     0
cinc   0   0   0   0   0  10   0   0   0   0     0
seis   0   0   0   0   0   0   9   1   0   0     0 [90.0/1.0]
siet   0   0   0   0   0   0   0  10   0   0     0
ocho   0   0   0   0   0   0   0   0  10   0     0
nuev   0   0   0   0   0   0   0   0   0  10     0
Ins    0   0   0   0   0   0   0   0   0   0
===================================================================
```

Figura 4.26: Análisis LPC con HTK

```
====================== HTK Results Analysis =======================
  Date: Sun Apr 07 09:45:00 2013
  Ref :
  Rec : rec.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=100.00 [H=100, S=0, N=100]
WORD: %Corr=100.00, Acc=100.00 [H=100, D=0, S=0, I=0, N=100]
------------------------ Confusion Matrix -------------------------
       c   u   d   t   c   c   s   s   o   n
       e   n   o   r   u   i   e   i   c   u
       r   o   s   e   a   n   i   e   h   e
       o           s   t   c   s   t   o   v
                       r   o       e       e   Del [ %c / %e]
cero  10   0   0   0   0   0   0   0   0   0     0
 uno   0  10   0   0   0   0   0   0   0   0     0
 dos   0   0  10   0   0   0   0   0   0   0     0
tres   0   0   0  10   0   0   0   0   0   0     0
cuat   0   0   0   0  10   0   0   0   0   0     0
cinc   0   0   0   0   0  10   0   0   0   0     0
seis   0   0   0   0   0   0  10   0   0   0     0
siet   0   0   0   0   0   0   0  10   0   0     0
ocho   0   0   0   0   0   0   0   0  10   0     0
nuev   0   0   0   0   0   0   0   0   0  10     0
Ins    0   0   0   0   0   0   0   0   0   0
===================================================================
```

Figura 4.27: Análisis MFCC con HTK



```
====================== HTK Results Analysis =======================
  Date: Sun Apr 07 08:10:56 2013
  Ref :
  Rec : rec.mlf
------------------------ Overall Results --------------------------
SENT: %Correct=100.00 [H=100, S=0, N=100]
WORD: %Corr=100.00, Acc=100.00 [H=100, D=0, S=0, I=0, N=100]
------------------------ Confusion Matrix -------------------------
       c   u   d   t   c   c   s   s   o   n
       e   n   o   r   u   i   e   i   c   u
       r   o   s   e   a   n   i   e   h   e
       o           s   t   c   s   t   o   v
                       r   o       e       e   Del [ %c / %e]
cero  10   0   0   0   0   0   0   0   0   0     0
 uno   0  10   0   0   0   0   0   0   0   0     0
 dos   0   0  10   0   0   0   0   0   0   0     0
tres   0   0   0  10   0   0   0   0   0   0     0
cuat   0   0   0   0  10   0   0   0   0   0     0
cinc   0   0   0   0   0  10   0   0   0   0     0
seis   0   0   0   0   0   0  10   0   0   0     0
siet   0   0   0   0   0   0   0  10   0   0     0
ocho   0   0   0   0   0   0   0   0  10   0     0
nuev   0   0   0   0   0   0   0   0   0  10     0
Ins    0   0   0   0   0   0   0   0   0   0
===================================================================
```

Figura 4.28: Análisis Filtros Cocleares con HTK

Capítulo 5

Conclusiones

La experiencia nunca se equivoca; es nuestra apreciación la que únicamente se equivoca, al esperar resultados no causados por los experimentos...
Leonardo da Vinci. La Verdadera Ciencia, Milán 1487.

5.1. Aportaciones científicas

En esta tesis se desarrolló una nueva solución del comportamiento de la membrana basilar aplicando análisis por resonancia al modelo del oído interno propuesto por Lesser y Berkeley [Les72], se considera de la solución compleja únicamente la expresión de la amplitud de la parte real y no se toma en cuenta el análisis de fase del sistema. Esta solución presenta la ventaja respecto a las soluciones existentes en la literatura de relacionar la frecuencia de excitación del sistema auditivo y la distancia en la cual se presenta la máxima amplitud de resonancia a lo largo de la membrana basilar, lo cual depende únicamente de las características físicas de masa, constante de elasticidad y resistencia mecánica a lo largo de la misma.

Esta solución fue validada en forma satisfactoria comparando los resultados obtenidos con los cuatro modelos más representativos del modelado del oído interno, los cuales son propuestos utilizando diferentes metodologías, teniendo el modelo de Peterson y Bogert [Pet50] su desarrollo con base en la integración numérica, el modelo de Lesser y Berkeley [Les72] una solución del comportamiento de la cóclea fundamentada en la mecánica de fluidos y el comportamiento de la membrana basilar como un sistema de osciladores armónicos forzados junto con el modelado por series de Fourier de la onda envolvente que se propaga sobre la membrana, el modelo de Allen [All77] considerando el empleo de la función de Green y su solución utilizando la ecuación de Laplace y por último el modelo de Neely [Nee81] el cual utiliza la aproximación por diferencias finitas para el modelado del comportamiento de la membrana basilar. En todos los casos los

resultados se evaluaron y compararon con el análisis por resonancia, obteniéndose concordancia total en los resultados obtenidos, enfatizando que la envolvente de la onda dada por la solucion de Lesser y Berkeley está en correspondencia con la forma de onda del análisis por resonancia.

Una validación adicional de la nueva solución del análisis por resonancia es realizada al comparar los resultados obtenidos con las observaciones experimentales de Békésy teniendo similitud en los resultados, lo cual resalta más el aporte científico que este trabajo brinda al estar en concordancia con resultados obtenidos de mediciones directas en la fisiología del oído interno y hace a la nueva solución completamente concordante con la teoría de los puntos de audición desarrollada por Békésy.

A partir de la nueva solución del análisis por resonancia, se desarrolló la representación en el dominio de la distancia de la membrana basilar a partir del espectro en frecuencia de una señal de voz lo cual fue llamado *membragrama*, siendo este planteamiento por primera vez presentado en este trabajo de Tesis. Esta nueva herramienta permite visualizar las zonas de mayor actividad a lo largo de la membrana basilar en presencia de los sonidos percibidos por el sistema auditivo. En la parte experimental de este trabajo de Tesis se utilizó para el análisis de la respuesta de la membrana basilar a los fonemas del español, pudiendo por primera vez observar la relación frecuencia distancia de su comportamiento. Esta herramienta puede ser utilizada en trabajos de investigación relacionados con la localización de posición de los electrodos sobre la membrana basilar en implantes cocleares.

Usando el análisis por resonancia se propone el desarrollo de un arreglo de filtros que modelan el comportamiento de la membrana basilar, teniendo la ventaja de poder estar delimitado por dos frecuencias dentro de las cuales se encuentra la parte de la señal de voz que presenta el mayor contenido de energía. Para la determinación de las frecuencias centrales se hace uso de la nueva solución haciendo una transformación del dominio de la frecuencia al dominio de la distancia para posteriormente segmentar la longitud definida de la cóclea utilizando el método de diferencias finitas y con los valores obtenidos determinar las frecuencias centrales del arreglo de filtros, concluyendo la extracción del vector paramétrico en forma similar a los MFCC. La metodología desarrollada se implementa en la herramienta HTK y se evalúa en procesos de reconocimiento de voz utilizando un corpus del idioma español obteniendo resultados satisfactorios.

5.2. Conclusiones

Se desarrolló un modelo mecánico acústico del oído interno usando el análisis por resonancia el cual modela en forma correcta el comportamiento físico de la membrana basilar considerando sus características físicas y proporcionando una relación unívoca entre la frecuencia de excitación del sistema auditivo y la distancia donde se presenta la máxima amplitud sobre la membrana basilar.

El modelo obtenido se comparó satisfactoriamente con los principales modelos existentes en la literatura y con las mediciones físicas reportadas del com-

portamiento coclear.

Se desarrolló una metodología de parametrización de la señal de voz mediante un arreglo de filtros triangulares diseñado a partir del análisis por resonancia los cuales modelan el comportamiento de la cóclea.

La parametrización desarrollada se evaluó en forma satisfactoria en procesos de reconocimiento de voz implementándola en HTK y comparándola con las metodologías existentes.

5.3. Productos obtenidos

Del trabajo desarrollado en esta Tesis se obtuvieron los siguientes productos:

- Simulation of a model of the basilar membrane in two dimensions for Spanish vowels. *160th Acoustical Society American Meeting. 15-16 Noviembre 2010, Cancún, México.*
- Modelado de la membrana basilar como un sistema de osciladores armónicos. *IV Congreso Nacional de Ingeniería en Comunicaciones y Electrónica. 24-26 Noviembre 2010, Ciudad de México, México.*
- Modelado del oído interno aplicado al análisis de señales de voz *19 Congreso Internacional Mexicano de Acústica. 5-7 Diciembre 2012, Ciudad de México, México.*
- Computational model of the cochlea using resonance analysis. *Revista Mexicana de Ingeniería Biomédica, Diciembre 2012, Volumen XXXIII, Número 2, páginas 77-86*
- Modelos del sistema auditivo aplicado a los sistemas de reconocimiento de voz. *Expoacústica, 28 Enero - 1 Febrero 2013. Ciudad de México, México.*

5.4. Trabajos futuros

La solución desarrollada usando análisis por resonancia es válida únicamente para los modelos en dos dimensiones de la cóclea, sin embargo en la literatura existen modelos en tres dimensiones del oído interno para los cuales la solución propuesta en esta tesis no es aplicable, los trabajos futuros deben de dar solución usando resonancia a este tipo de modelos.

El análisis por resonancia desarrollado no contempla la relación entre la velocidad de propagación de la onda dentro de la endolinfa y su relación con el cambio de fase de la onda, debido a que únicamente considera áreas específicas sobre la membrana basilar con un comportamiento equivalente al de un oscilador armónico forzado amortiguado, sin embargo los trabajos futuros con análisis similares tienen que modelar esta propiedad.

Debido a que la solución desarrollada establece que la fuerza de excitación está normalizada, no proporciona el valor de la presión ejercida sobre la membrana basilar para cada valor de distancia a lo largo de la cóclea, los modelos futuros basados en este tipo de soluciones deben de proporcionar esta relación.

La solución por resonancia únicamente considera las características físicas de la membrana basilar y por lo tanto no es posible determinar el factor de atenuación dentro de la perilinfa, lo cual debe ser resuelto en trabajos posteriores basados en la mecánica de fluidos de la cóclea.

En la literatura del modelado mecánico acústico del oído interno pocos modelos contemplan la geometría en espiral de este órgano, es necesario que las soluciones posteriores modelen esta característica física y proporcionen su funcionalidad.

Los trabajos futuros deben considerar en forma conjunta los modelos fisiológicos y de procesamiento de señales de la cóclea junto con los modelos físicos del oído interno, para proponer una metodología para la obtención de parámetros cocleares que no solo contemplen la respuesta mecánica de la membrana basilar, sino que también contemplen la respuesta fisiológica del órgano de Corti y del nervio auditivo.

Referencias Bibliográficas

[All77] Allen J. B. *Two-dimensional cochlear fluid model: New Results*, JASA, Vol. 61, pp. 110-119, 1977.

[All79] Allen J. B., Sondhi M. M. *Cochlear macromechanics: Time domain solutions*, JASA, Vol. 66, pp. 123-132, 1979.

[All85] Allen J. B. *Cochlear Modelling*, IEEE ASSP Magazine, pp. 3-29, 1985.

[Alo67] Alonso M., Finn E. J. *University Physics Mechanics*, Addiosn-Wesley Publishing Company, USA, 1967.

[Ata71] Atal B. S., Hanauer S. L. *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave*, JASA, Vol. 50, pp. 637-655, 1971.

[Bah83] Bahl L. R., Jelinek F., Mercer R. L. *A Maximum Likelihood Approach to Continuous Speech Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-5, pp. 179-190, 1983.

[Bak75] Baker J. K. *The DRAGON System-An Overview*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, pp. 24-29, 1975.

[Bat67] Batchelor G. K. *An Introduction to Fluid Dynamics*, Cambridge University Press, Inglaterra, 1967.

[Bek60] Békésy V. v. *Experiments in Hearing*, Mc Graw Hill, USA, 1960.

[Ben08] Benesty J., Sondhi M. M., Huang Y. *Springer Handbook of Speech Processing*, Springer, USA, 2008.

[Ber00] Bernal B. J., Bobadilla S. J., Gómez V. P. *Reconocimiento de Voz y Fonética Acústica*, RA-MA, España, 2000.

[Boa06] Boas M. L. *Mathematical Methods in the Physical Sciences*, John Wiley & Sons, Inc., Tercera Edición, USA, 2006.

[Boe80] de Boer E. *Auditory Physics. Physical principles in hearing I*, Phys. Rep., Vol. 62, pp. 97-274, 1980.

[Boe84] de Boer E. *Auditory Physics. Physical principles in hearing II*, Phys. Rep., Vol. 105, pp. 141-226, 1984.

[Boe96] de Boer S. E., Ed. by: Dallos P., Fay R. R. *The cochlea Chapter 5: Mechanics of the cochlea: Modeling effects*, Springer, USA, 1996.

[Bog63] Bogert B. P., Healy M. J. R., Tukey J. W. *The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking*, Proceedings of the Symposium on Time Series Analysis, Wiley, USA, 1963.

[Cas89] Casacuberta N. F., Vidal R. E. *Reconocimiento automático del habla*, Marcombo, España, 1987.

[Cha80] Chadwick R. *Studies in cochlear mechanics*, Mathematical Modeling of the Hearing Process Lecture Notes in Biomathematics, Springer, 1980.

[Cha02] Chapra C. S., Canale P. R. *Numerical Methods for Engineers With Software and Programming Aplications*, Cuarta Edición, Mc Graw Hill, USA, 2002.

[Che88] Chen C. H. *Signal Processing Handbook*, Marcel Dekker, Inc., USA, 1988.

[Coh95] Cohen L. *Time-Frequency Analysis*, Prentice Hall, USA, 1995.

[Coo65] Cooley. J. W., Tukey J. W. *An Algorithm for the Machine Calculation of Complex Fourier Series*, Math Computation, Vol. 19, pp. 297-301, 1965.

[Dal90] Dallos P., Geisler C. D., Matthews J. W., Ruggero M. A., Steele C. R. *The Mechanics and Biophysics of Hearing*, Lecture Notes in Biomathematics, Vol. 87, Springer, 1990.

[Dal96] Dallos P., Popper A. N., Fay R. R. *The Cochlea*, Springer, USA, 1996.

[Dav83] Davis H. *An active process in cochclear mechanics*, Hear. Res., Vol. 9, pp. 1-49, 1983.

[Dav80] Davis S. B., Mermelstein P. *Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-28, pp. 357-366, 1980.

[Del55] Delattre P. C., Liberman A. M., Cooper F. S. *Acoustic Loci and Transitional Cues for Consonants*, JASA, Vol. 27, pp. 769-773, 1955.

[Del00] Deller J. R. Jr., Hansen J. H. L., Proakis J. G. *Discrete-Time Processing of Speech Signals*, IEEE Wiley-Interscience, USA, 2000.

[Den85] Den Hartog J. P. *Mechanical Vibrations*, Dover, USA, 1985.

[Dud39] Dudley H. *Remaking Speech*, JASA, Vol. 11, pp. 169-177, 1939.

[Dun61] Dunn H. K. *Methods of Measuring Vowel Formant Bandwidths*, JASA, Vol. 33, pp. 1737-1746, 1961.

[Eat00] Eatock R. A. *Adaptation in hair cells*, Annual Review of Neuroscience, Vol. 23, pp. 285-314, 2000.

[Eco93] Eco U. *La Ricerca della Lingua Perfetta nella Cultura Europea*, Lazerta, Italia, 1993.

[Ell07] Elliott S. J., Ku E. M., Lineton B. *A state space model for cochlear mechanics*, JASA, Vol. 122, pp. 2759-2771, 2007.

[Ell11] Elliott S. J., Lineton B., Ni G. *Fluid coupling in a discrete model of cochlear mechanics*, JASA, Vol. 130, pp. 1441-1451, 2011.

[Elm85] Elmore W. C., Heald M. A. *Physics of Waves*, Dover, USA, 1985.

[Fau01] Faúndez Z. M. *Tratamiento Digital de Voz e Imagen*, Marcombo, España, 2001.

[Fla68] Flanagan J. L., Landgraf L. L. *Self-Oscillating Source for Vocal-Tract Synthesizers*, IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, pp. 57-64, 1968.

[Fla69] Flanagan J. L., Cherry L. *Excitation of Vocal-Tract Synthesizers*, JASA, Vol. 45, pp. 764-769, 1969.

[Fla70] Flanagan J. L., Coker C. H., Rabiner L. R., Schafer R. W., Umeda N. *Synthetic Voices for Computers*, IEEE Spectrum, pp. 22-45, 1970.

[Fla76] Flanagan J. L. *Computers that Talk and Listen: Man-Machine Communication by Voice*, Proceedings of the IEEE, Vol. 64, pp. 405-415, 1976.

[Fle33] Fletcher H., Munson W. A. *Loudness, Its Definition, Measurement and Calculation*, JASA, Vol. 5, pp. 82-108, 1933.

[Fle51] Fletcher H. *On the dynamics of the cochlea*, JASA, Vol. 23, pp. 637-645, 1951.

[Fuj62] Fujimura O. *Analysis of Nasal Consonants*, JASA, Vol. 34, pp. 1865-1875, 1962.

[Fur92] Furui S. Sondhi M. M. *Advances in Speech Signal Processing*, Marcel Dekker, Inc., USA, 1993.

[Fur01] Furui S. *Digital Speech Processing, Synthesis, and Recognition*, Segunda Edición, Marcel Dekker, Inc., USA, 2001.

[Gei76] Geisler C. D. *Mathematical Models of the Mechanics of the Cochlea*, Handbook of Sensory Physiology, Springer, USA, 1976.

[Ger83] Gersho A., Cuperman V. *Vector Quantization: A Pattern-Matching Technique for Speech Coding*, IEEE Communications Magazine, pp. 15-21, 1983.

[Gol68] Gold B., Rabiner L. R. *Analysis of Digital and Analog Formant Synthesizers*, IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, pp. 81-94, 1968.

[Gol11] Gold B., Morgan N., Ellis D. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, Wiley, USA, 2011.

[Gra84] Gray R. M. *Vector Quantization*, IEEE ASSP Magazine, pp. 4-29, 1984.

[Hay99] Hayes M. H. *Digital Signal Processing*, Mc Graw Hill, USA, 1999.

[Hei61] Heinz J. M., Stevens K. N. *On the Propierties of Voiceless Fricative Consonants*, JASA, Vol. 33, pp. 589-596, 1961.

[Hel54] Helmholtz H. L. F. *On the Sensations of Tone as a Physiological basis for the Theory of Music*, Dover, USA, 1954.

[Hel67] Helms H. D. *Fast Fourier Transform Method of Computing Difference Equations and Simulating Filters*, IEEE Transactions on Audio and Electroacoustics, Vol. AU-15, pp. 85-90, 1967.

[Her90] Hermansky H. *Perceptual linear predictive (PLP) analysis of speech*, JASA, Vol. 87, pp. 1738-1752, 1990.

[Hil00] Hilera R. J., Martínez J. V. *Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones*, Ra-Ma, España, 2000.

[Hol80a] Holmes M. H. *An analysis of a low-frequency model of the cochlea*, JASA, Vol. 68, pp. 482-488, 1980.

[Hol80b] Holmes M. H. *Low frequency asymptotics for a hydroelastic model of the cochlea*, SIAM Journal on Applied Mathematics, Vol. 38, pp. 445-456, 1980

[Hol82] Holmes M. H. *A Mathematical model of the dynamics of the inner ear*, Journal of Fluid Mechanics, Vol. 116, pp. 59-75, 1982.

[Hud85] Hudspeth A. J. *The cellular basis of hearing: the biophysics of hair cells*, Science, Vol. 230, pp. 745-752, 1985.

[Ins76] Inselberg A., Chadwick R. S. *Mathematical model of the cochlea. I: formulation and solution*, SIAM Journal of Applied Mathematics, Vol. 30, pp. 149-163, 1976.

[Iso90] Iso K., Watanabe T. *Speaker-Independent Word Recognition using a Neural Prediction Model*, IEEE CH2847, pp. 441-444, 1990.

[Ita75] Itakura F. *Minimum Prediction Residual Principle Applied to Speech Recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-23, pp. 67-72, 1975.

[Jim10a] Jiménez Hernández M., Oropeza Rodríguez. J. L., Suárez Guerra S. *Simulation of a model of the basilar membrane in two dimensions for Spanish vowels*, 160th Acoustical Society of America Meeting, México, 2010.

[Jim10b] Jiménez Hernández Mario. *Modelado de la membrana basilar como un sistema de osciladores armónicos*, IV Congreso Nacional de Ingeniería en Comunicaciones y Electrónica, México, 2010.

[Jim12] Jiménez Hernández M., Oropeza Rodríguez. J. L., Suárez Guerra. S., Barrón Fernández. R. *Computational Model of the Cochlea using Resonance Analysis*, Revista Mexicana de Ingeniería Biomédica, Vol. 33, pp. 77-86, 2012.

[Jim13] Jiménez Hernández M., Oropeza Rodríguez. J. L., Suárez Guerra. S., Barrón Fernández. R. *Resonance analysis of the cochlea applied to speech recognition*, En preparación, Applied Acoustics, 2013.

[Kal97] Kalouptsidis N. *Signal Processing Systems Theory and Design*, Wiley-Interscience, USA, 1997.

[Kee98] Keener J. P. *Principles of Applied Mathematics, Transformation and Approximation*, Segunda Edición, Perseus Books, Inglaterra, 1998.

[Kee08] Keener J., Sneyd J. *Mathematical Physiology*, Segunda Edición, Springer, USA, 2008.

[Kev96] Kevorkian J., Cole J. D. *Multimple Scale and Singular Perturbation Methods*, Springer, USA, 1996.

[Kim80] Kim D. O., Neely S. T., Molnar C. E., Matthews J. W. *An active cochlear model with negative damping in the cochlear partition: Comparison with Rhode's ante-and post-mortem results*, Physicological, Physiological, and Behavioral Studies in Hearing, Delf U. P., Holanda, 1980.

[Kin00] Kinsler L. E., Frey A. R., Coppens A. B., Sanders J. V. *Fundamentals of Acoustics*, Cuarta Edición, John Wiley & Sons, Inc., USA, 2000.

[Koh88] Kohonen T. *The "Neural" Phonetic Typewriter*, IEEE Computer, pp. 11-22, 1988.

[Ku08] Ku E. M., Elliott S. J., Lineton B. *Statistics of instabilities in a state space model of the human cochlea*, JASA, Vol. 124, pp. 1068-1069, 2008.

[Les72] Lesser M. B., Berkley D. A. *Fluid mechanics of the cochlea. Part 1*, J. Fluid Mech., Vol. 51, pp. 497-512, 1972.

[Lie73] Lien M. D. *A Mathematical Model of the Mechanics of the cochlea*, Tesis Doctoral, Universidad de Washington, USA, 1973.

[Lin80] Linde Y., Buzo A., Gray R. M. *An Algorithm for Vector Quantizer Design*, IEEE Transactions on Communications, Vol. Com-28, pp. 84-95, 1980.

[Lin03] Lindgren A. G., Li W. *Analysis and simulation of a classic model of cohlea mechanics via a state-space realization*, Reporte Técnico, Universidad de Rhode Island, USA, 2003.

[Mar07] Martín B. B., Sanz M. A. *Redes Neuronales y Sistemas Borrosos*, Tercera Edición, Ra-Ma, España, 2007.

[McD91] McDermott E., Katagiri S. *LVQ-Based Shift-Tolerant Phoneme Recognition*, IEEE Transactions on Signal Processing, Vol. 39, pp. 1398-1411, 1991.

[Mea91] Meade M. L., Dillon C. R. *Signals and Systems: Models and Behaviour*, Segunda Edición, Chapman & Hall, Inglaterra, 1991.

[Mon88] Monderer B., Lazar A. A. *Speech signal detection at the output of a cochlear model*, ICASSP, Vol. 14, pp. 11-14, 1988.

[Nee81] Neely S. T. *Finite difference solution of a two-dimensional mathematical model of the cochlea*, JASA, Vol. 69, pp. 1386-1393, 1981.

[Nee83] Neely S. T., Kim D. O. *An active cochlear model showing sharp tuning and high sensitivity*, Hear. Res., Vol. 9, pp. 123-130, 1983.

[Nee85] Neely S. T. *Mathematical modeling of cochlear mechanics*, JASA, Vol.78, pp. 345-352, 1985.

[Nee86] Neely S. T., Kim D. O. *A model for active elements in cochlear biomechanics*, JASA, Vol. 79, pp. 1472-1480, 1986.

[Nol64] Noll A. M. *Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection*, JASA, Vol. 36, pp. 296-302,1964.

[Nol67] Noll A. M. *Cepstrum Pitch Determination*, JASA, Vol. 41, pp. 293-309, 1967.

[Nyq28] Nyquist H. *Certain Topics in Telegraph Transmission Theory*, Transactions of the AIEE, pp. 617-644, Febrero 1928.

[Opp68] Oppenheim A. V., Schafer R. W. *Homomorphic Analysis of Speech*, IEEE Transactions on Audio and Electroacoustics, Vol. AU-16, pp. 221-226, 1968.

[Opp69] Oppenheim A. V. *A Speech Analysis-Synthesis System Based on Homomorphic Filtering*, JASA, Vol. 45, pp. 458-465, 1969.

[Opp75] Oppenheim A. V., Schafer R. W. *Digital Signal Processing*, Prentice Hall, USA, 1975.

[Opp97] Oppenheim A. V., Willsky A. S., Nawab S. H. *Signals & Systems*, Segunda Edición, Prentice Hall, USA, 1997.

[Opp99] Oppenheim A. V., Schafer R. W., Buck J. R. *Discrete-Time Signal Processing*, Segunda Edición, Prentice Hall, USA, 1999.

[Pes76] Peskin C. S. *Partial Differential Equations in Biology*, Sciences Lectures Notes, USA, 1976.

[Pes81] Peskin C. S. *Lectures on mathematical aspects of physiology*, AMS Lectures in Applied Mathematics, Vol. 19, pp. 38-69, 1981.

[Pet50] Peterson L. C., Bogert B. P. *A Dynamical Theory of the Cochlea*, JASA, Vol. 22, pp. 369-381, 1950.

[Pet52] Peterson G. E., Barney H. L. *Control Methods Used in a Study of the Vowels*, JASA, Vol. 24, pp. 175-184, 1952.

[Por73] Portnoff M. R., Schafer R. W. *Mathematical Considerations in Digital Simulations of the Vocalt Tract*, JASA, Vol. 53, pp. 294, 1973.

[Pro96] Proakis J. G., Manolakis D. G. *Digital Signal Processing Principles, Algorithms, and Applications*, Tercera Edición, Prentice Hall, USA, 1996.

[Pro04] Proakis J. G., Ingle V. K. *A Self-Study Guide for Digital Signal Processing*, Prentice Hall, USA, 2004.

[Qua02] Quatieri T. F. *Discrete-Time Speech Signal Processing Principles and Practice*, Prentice Hall, USA, 2002.

[Qui99] Quilis A. *Tratado de Fonología y Fonética Españolas*, Segunda Edición, Gredos, España, 1999.

[Rab68] Rabiner L. R. *Digital-Formant Synthesizer for Speech-Synthesis Studies*, JASA, Vol. 43, pp. 822-828, 1968.

[Rab75] Rabiner L. R., Gold B. *Theory and Applications of Digital Signal Processing*, Prentice Hall, USA, 1975.

[Rab76] Rabiner L. R., Schafer R. W. *Digital Techniques for Computer Voice Response: Implementations and Applications*, Proceedings of the IEEE, Vol. 64, pp. 416-433, 1976.

[Rab78] Rabiner L. R., Schafer R. W. *Digital Processing of Speech Signals*, Prentice Hall, USA, 1978.

[Rab79] Rabiner L. R., Levinson S. E., Rosenberg A. E., Wilpon J. G. *Speaker - Independent Recognition of Isolated Words Using Clustering Techniques*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, pp. 336-349, 1979.

[Rab81] Rabiner L. R., Levinson S. E. *Isolated and Connected Word Recognition - Theory and Selected Applications*, IEEE Transactions on Communications, Vol. COM-29, No. 5, pp. 621-659, 1981.

[Rab89a] Rabiner L. R., Wilpon J. G., Soong F. K. *High Performance Connected Digit Recognition Using Hidden Markov Models*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, pp. 1214-1225, 1989.

[Rab89b] Rabiner L. R. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of the IEEE, Vol. 77, pp. 257-286, 1989.

[Rab93] Rabiner L. R., Juang B. H. *Fundamentals of Speech Recognition*, Prentice Hall, USA, 1993.

[Ran50] Ranke O. F. *Theory of Operation of the Cochlea: A Contribution to the Hidrodinamycs of the Cochlea*, JASA, Vol. 22, pp. 772-777, 1950.

[Red76] Reddy D. R. *Speech Recognition by Machine: A review*, Proceedings of the IEEE, Vol. 64, pp. 501-531, 1976.

[Rho71] Rhode W. S. *Observations of the Vibration of the Basilar Membrane in Squirrel Monkeys using the Mössbauer Technique*, JASA, Vol. 49, pp. 1218-1231, 1971.

[Rho74] Rhode W. S., Robles L. *Evidence of Mossbauer experiments for nonlinear vibrarion in the cochlea*, JASA, Vol. 55, pp. 558-596, 1974.

[Rho84] Rhode W. S. *Cochlear mechanics*, Annual Review of Physiology, Vol. 46, pp. 231-264, 1984.

[Ros71] Rosenberg A. E. *Effect of Glottal Pulse Shape on the Quality of Natural Vowels*, JASA, Vol. 49, pp. 583-590, 1971.

[Ros07] Rossing T. D. *Springer Handbook of Acoustics*, Springer, USA, 2007.

[Rou87] Roucos S., Dunham M. O. *A Stochastic Segment Model for Phoneme-Based Continuous Speech Recognition*, IEEE CH2396, pp. 73-76, 1987.

[Sak78] Sakoe H., Chiba S. *Dynamic Programming Algorithm Optimization for Spoken Word Recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, pp. 43-49, 1978.

[Sak79] Sakoe H. *Two-Level DP-Matching-A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, pp. 588-595, 1979.

[Sak89] Sakoe H., Isotani R., Yoshida K., Iso K., Watanabe T. *Speaker-Independent Word Recognition Using Dynamic Programming Neural Networks*, IEEE CH2673, pp. 29-32, 1989.

[Sch75] Schafer R. W., Rabiner L. R. *Digital Representations of Speech Signals*, Proceedings of the IEEE, Vol. 63, pp. 662-677, 1975.

[Sch88] Scheid F. *Numerical Analysis*, Segunda Edición, Mc Graw Hill, USA, 1988.

[Set71] Seto W. W. *Theory and Problems of Acoustics*, McGraw-Hill Book Company, USA, 1971.

[Sha48] Shannon C. E. *A Mathematical Theory of Communication*, The Bell System Thecnical Journal, Vol. 27, pp.379-423, 1948.

[Sie74] Siebert W. M. *Ranke Reviseted-a simple short-wave cochlear model*, JASA, Vol. 56, pp. 594-600, 1974.

[Son71] Sondhi M. M., Gopinath B. *Determination of Vocal-Tract Shape from Impulse Response at the lips*, JASA, Vol. 49, pp. 1867-1873, 1971.

[Son74] Sondhi M. M. *Model for wave propagation in a lossy vocal tract*, JASA, Vol. 55, pp. 1070-1075, 1974.

[Ste74] Steele C. R. *Behavior of the basilar membrane with pure-tone excitation*, JASA, Vol. 55, pp. 148-162, 1974.

[Ste79] Steele C. R., Taber L. A. *Comparison of WKB calulations and experimental results for three-dimensional cochlear models*, JASA, Vol. 65, pp. 1007-1018, 1979.

[Sti98] Stibler B. Z., Lewis E. R., Henry K. R. *A state space model of gerbil cochlea*, Proceedings of the 6th Annual Conference on Computational Neuroscience: Trends in research, pp. 107-112, 1998.

[Vie75] Viergever M. A., Kalker J. J. *A two dimensional model for the cochlea I. The exact approach*, J. Eng. Math., Vol. 9, pp. 353-365, 1975.

[Vie77] Viergever M. A. *A two dimensional model for the cochlea II. The heuristic approach and numerical results*, J. Eng. Math., Vol. 11, pp. 11-28, 1977.

[Vie80] Viergever M. A. *Mechanics of the Inner Ear - A Mathematical Approach*, Delft U. P., Holanda, 1980.

[Wai89a] Waibel A., Hanazawa T., Hinton G., Shikano K., Lang K. J. *Phoneme Recognition Using Time-Delay Neural Networks*, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, pp. 328-339, 1989.

[Wai89b] Waibel A., Sawai H., Shikano K. *Consonant Recognition by Modular Construction of Large Phonemic Time-Delay Neural Networks*, IEEE CH2673, pp. 112-115, 1989.

[Wai90] Waibel A., Lee K. F. *Readings in Speech Recognition*, Morgan Kaufmann Publishers, Inc., USA, 1990.

[Wyl75] Wylie C. R. *Advanced Engineering Mathematics*, McGraw-Hill Book Company, Cuarta Edición, USA, 1975.

[Win70] Winham G., Steiglitz K. *Input Generators for Digital Sound Synthesis*, JASA, Vol. 47, pp. 665-666, 1970.

[Yos06] Yost W. A. *Fundamentals of Hearing An Introduction*, Quinta Edición, Academic Press, USA, 2006.

[You09] Young S., Evermann G., Gales M., Hain T., Kershaw D., Liu X., Moore G., Odell J., Ollason D., Povey D., Valtchev V., Woodland P. *The HTK Book*, Cambridge University Engineering Department, Inglaterra, 2009.

[Zwi50] Zwislocki J. *Theory of the Acoustical Action of the Cochlea*, JASA, Vol. 22, pp. 778-784, 1950.

[Zwi53] Zwislocki J. *Review of Recent Mathematical Theories of Cochlear Dynamics*, JASA, Vol. 25, pp. 743-751, 1953.

Apéndice A

Solución numérica de la mecánica de fluidos en la cóclea por Lesser y Berkeley

A partir de las condiciones límites se puede considerar la simetría física del modelo utilizando únicamente las soluciones impares en el eje y, considerando sólo la región superior del modelo. Cuando la entrada tiene una sola frecuencia $F(y,t) = \widehat{F}(y)e^{j\omega t}$ entonces $\phi(x,y,t)$ es de la forma $\widehat{\phi}(x,y,\omega)e^{j\omega t}$ y similarmente para las otras variables, planteando las soluciones de esta forma para todas las variables se obtienen las siguientes ecuaciones.

$$\nabla^2\widehat{\phi} = 0, \quad \widehat{p} + j\omega\rho\widehat{\phi} = 0 \tag{A.1}$$

$$\frac{\partial\widehat{\phi}}{\partial y} = j\omega\widehat{\eta}, \quad j\omega\widehat{\eta}Z = -\widehat{p}, \quad y = 0 \tag{A.2}$$

$$\frac{\partial\widehat{\phi}}{\partial x} = U_0, \quad x = 0 \tag{A.3}$$

$$\frac{\partial\widehat{\phi}}{\partial x} = 0, \quad x = L \tag{A.4}$$

$$\frac{\partial\widehat{\phi}}{\partial y} = 0, \quad y = l \tag{A.5}$$

Donde $Z = j\omega m + R_m + k/(j\omega)$ es la impedancia del oscilador armónico forzado y $U_0 = j\omega\widehat{F}$, para obtener las soluciones en el dominio de la frecuencia las ecuaciones diferenciales se transforman en ecuaciones algebraicas. Se considera además que en la ecuación A.2 la presión es una función impar de y. Es necesario realizar un ajuste dimensional normalizando a las ecuaciones del

modelo por medio del escalamiento de las variables x y y por L, Z por $j\omega\rho L$, y $\widehat{\phi}$ por $U_0 L$, una vez realizado el reacomodo de variables se obtiene el siguiente conjunto de ecuaciones.

$$\nabla^2\phi = 0 \tag{A.6}$$

$$\frac{\partial\phi}{\partial y} = \frac{2\phi}{Z}, \quad y = 0 \tag{A.7}$$

$$\frac{\partial\phi}{\partial x} = 1, \quad x = 0 \tag{A.8}$$

$$\frac{\partial\phi}{\partial x} = 0, \quad x = 1 \tag{A.9}$$

$$\frac{\partial\phi}{\partial y} = 0, \quad y = \sigma \tag{A.10}$$

Donde $\sigma = l/L$, una solución analítica para este problema se obtiene utilizando series de Fourier de la siguiente forma.

$$\phi = x(1 - \frac{x}{2}) - \sigma y(1 - \frac{y}{2\sigma}) + \sum_{n=0}^{\infty} A_n cosh[n\pi(\sigma - y)]cos(n\pi x) \tag{A.11}$$

Siendo en la ecuación la constante A_n desconocida. Debido a que ϕ satisface todas las condiciones límites excepto la planteada por la ecuación A.7 se puede utilizar a esta ecuación para determinar los coeficientes desconocidos A_n, teniendo entonces la siguiente expresión.

$$\sigma + \sum_{n=0}^{\infty} n\pi A_n sinh(n\pi\sigma)cos(n\pi x) - \frac{2}{Z}[x(1 - x/2) + \sum_{n=0}^{\infty} A_n cosh(n\pi\sigma)cos(n\pi x)] = 0 \tag{A.12}$$

Truncando la serie en N términos, multiplicándola por $cos(m\pi x)$, e integrando desde 0 hasta 1, se obtiene el siguiente sistema de ecuaciones lineales.

$$\sum_{n=0}^{N} A_n \alpha_{nm} = f_m \tag{A.13}$$

Donde α_{nm} está dada de la siguiente forma.

$$\alpha_{mn} = 2cosh(n\pi\sigma) \int_0^1 \frac{cos(n\pi x)cos(m\pi x)}{Z} dx - \frac{1}{2} n\pi sinh(n\pi\sigma)\delta_{nm} \tag{A.14}$$

Donde f_m se obtiene de la siguiente expresión.

$$f_m = \sigma\delta_{m0} - \int_0^1 \frac{x(2-x)cos(m\pi x)}{Z}dx \tag{A.15}$$

Siendo $\delta_{ij} = 1$ si $i = j$ y teniendo el valor de 0 bajo cualquier otra circunstancia. Debido a que f_m y α_{mn} pueden ser evaluadas en forma explícita, ésto permite poder obtener un conjunto de N ecuaciones lineales para A_n en el intervalo $1 \leq n \leq N$, lo cual cuando $N \to \infty$ da la solución de las ecuaciones del modelo cuando se sustituyen en la ecuación A.11.

Apéndice B

Fundamentos de las metodologías de reconocimiento de voz

B.1. DTW

El alineamiento temporal dinámico (DTW) consiste en ajustar la posición de los fonemas entre la señal de entrada al reconocedor y las estructuras de referencia del sistema mediante la expansión y compresión del eje del tiempo. Este proceso se hace más eficiente al utilizar la programación dinámica y se utiliza principalmente para aplicaciones no robustas. Esta metodología consiste en la comparación de secuencias de dos vectores característicos en el dominio del tiempo de la señal de voz [Fur01].

$$A = a_1, a_2, ..., a_i \quad y \quad B = b_1, b_2, ..., b_j \tag{B.1}$$

Una vez obtenidos los vectores se considera un plano definido por los vectores A y B en donde la función de alineamiento en el tiempo está indicada por la correspondencia entre los ejes del tiempo de las secuencias A y B, las cuales pueden ser representadas por una secuencia de puntos sobre el plano $c = (i, j)$ de la siguiente forma.

$$F = c_1, c_2, ..., c_k, c_K \quad C_k = (i_k, j_k) \tag{B.2}$$

Donde la distancia entre los dos vectores característicos a_i y b_j se representa por $d(c) = d(i, j)$, la suma de las distancias desde el inicio hasta el final de las secuencias a lo largo de F puede ser representada por:

$$D(F) = \frac{\sum_{k=1}^{K} d(c_k) w_k}{\sum_{k=1}^{K} w_k} \tag{B.3}$$

El valor más pequeño de la función es el que mejor se aproxima a los dos vectores A y B, siendo la función w_k una función de peso positiva cuyo valor esta relacionado a la función F. La minimización que se realiza en la ecuación anterior debe cumplir tres condiciones, la de monotonía y continuidad, dada por:

$$0 \leq i_k - i_{k-1} \leq 1 \qquad 0 \leq j_k - j_{k-1} \leq 1 \tag{B.4}$$

la condición de frontera:

$$i_1 = j_1 = 1 \qquad i_K = I, j_K = J \tag{B.5}$$

y la condición de ajuste de ventana:

$$|i_k - j_k| \leq r \qquad r = constante \tag{B.6}$$

La última condición se aplica para prevenir la expansión y contracción extrema, definiendo a w_k tal que el denominador de la ecuación 2.54 sea una constante independiente de F pudiendo simplificar la ecuación.

B.2. VQ

La metodología de cuantificación vectorial (VQ) consiste en mapear un vector x con respecto a otro vector k-dimensional y, estando x cuantificada como y de la forma [Fur01]:

$$y = q(x) \tag{B.7}$$

El conjunto de valores y es finito de la forma $Y = \{y_i\}$ $(1 \leq i \leq K)$, el conjunto Y es referido a un libro código y el arreglo $\{y_i\}$ son los vectores código. El tamaño del libro código K es una referencia del número de niveles. Para el diseño del libro código el espacio k-dimensional del vector x es particionado en K regiones $\{c_i\}$ $(1 \leq i \leq K)$ con un vector y_i asociado con cada una de las regiones C_i lo cual se representa como:

$$q(x) = y_i \tag{B.8}$$

Cuando x es cuantificada como y, una medida de distorsión de la cuantificación o medida de distancia $d(x, y)$ puede ser definida entre x y y, la distorción promedio total está dada entonces por:

$$D = lim_{M\to\infty} \frac{1}{M} \sum_{n=1} Md[x(n), y(n)] \tag{B.9}$$

Se considera que un cuantificador es óptimo si la distorsión total es mínima en todos los nivles K cuantificados, existen dos condiciones necesarias para su optimización, la primera es que el cuantificador se diseñe utilizando la mínima distorsión o regla de selección del vecino más próximo, la cual se expresa de la forma:

$$q(x) = y_i \quad si \quad d(x, y_i) \leq d(x, y_j) \qquad j \neq i \quad 1 \leq j \leq K \tag{B.10}$$

La segunda condición es que cada uno de los vectores código y_i se escoja para minimizar la distorsión promedio en la región C_i, a tal vector se le denomina centroide de la región C_i. El centroide para una región particular depende de la definición de la medida de distorsión.

B.3. HMM

Un modelo de Markov oculto (HMM) es un modelo que puede ser descrito en cualquier instante de tiempo como un conjunto de N estados dados por $\{1, 2, ..., N\}$, en tiempos discretos y espacios regulares. El sistema cambia de estado de acuerdo a un conjunto de posibilidades asociadas con cada estado, los instantes de tiempo asociados con los cambios de estado son $t = 1, 2, ...$ y el estado actual de cada tiempo t se denota como q_t, una descripción probabilística actual del sistema requiere la especificación de los estados actuales en el tiempo t y los actuales predecedores, para el caso especial de un tiempo discreto de primer orden, la dependencia probabilística es truncada a solo un estado predecedor, lo cual está dado por [Rab93]:

$$P[q_t = j | q_{t-1} = i, q_{t-2} = k, ...] = P[q_t = j | q_{t-1=i}] \tag{B.11}$$

Se consideran únicamente los procesos en los cuales los estados a la derecha del sistema son independientes del tiempo, estableciendo el conjunto de probabilidades de transición de estados a_{ij} de la forma siguiente:

$$a_{ij} = P[q_t = |q_{t-1} = i] \qquad 1 \leq i, j \leq N \tag{B.12}$$

Para lo cual se tienen las siguientes propiedades:

$$a_{ij} \geq 0 \qquad \forall j, i \tag{B.13}$$

$$\sum_{j=1}^{N} a_{ij} = 1 \qquad \forall i \tag{B.14}$$

El proceso estocástico es entonces un modelo de Markov observable teniendo a la salida del proceso el conjunto de estados para cada instante de tiempo, donde cada estado corresponde a un evento observable.

B.4. NN

Una red neuronal (NN) está compuesta por varios nodos o elementos no lineales simples que operan en paralelo simulando un patrón biológico de redes neuronales, cada nodo está caracterizado por un umbral interno θ y por una función de activación, existen tres tipos de funciones de activación, de tipo

escalón, tipo lineal mixta y sigmodal. Los modelos más comunes de NN son los perceptrones multicapa los cuales tienen uno o más nodos a la entrada y varios nodos a la salida, su funcionamiento se basa en una regla de desición dependiendo de los nodos de salida [Fur01].

Para un perceptron x_i' y x_k'' son las salidas de los nodos en la primera y segunda capa oculta, θ_i' y θ_k'' son los umbrales internos en estos nodos, w_{ij} es la conexión desde la entrada hasta la primera capa oculta, ij' es la conexión entre la primera y la segunda capa y w_{ij}' es la conexión entre la segunda capa y la capa de salida, se considera que el sistema tiene una función de activación $f(\alpha)$ que está dada en su forma general como:

$$f(\alpha) = \frac{1}{1 + e^{-(\alpha - \theta)}} \tag{B.15}$$

La toma de desiciones requiere de un algoritmo de clasificación que pueda ser generado a partir de las tres capas, permitiendo a los perceptrones ser entrenados automáticamente mediante un algoritmo de back-propagation, este tipo de algoritmo minimiza el error cuadrático medio entre la salida actual de la red y la salida deseada, permitiendo al sistema trabajar como un clasificador al obtener a la salida los valores de 1 o 0 dependiendo de los valores de entrada de la red, la forma de entrenar un algoritmo back-propagation está divida en cuatro etapas.

En la primera etapa se hace la inicialización del umbral y el peso, colocando todos los pesos y los umbrales de los nodos con valores aleatorios pequeños, en la etapa dos se realiza la entrada y presentación de la salida deseada, considerando como entrada un vector de valores continuos x_0, x_1, x_{N-1}, y especificando las salidas deseadas d_0, d_1, d_{M-1}, presentando las muestras de un entrenamiento en forma cíclica hasta estabilizar los pesos, el paso tres es el cálculo de la salida actual, el cual se logra usando la no linealidad y el cálculo de las salidas y_0, y_1, y_{M-1}, una vez realizado ésto, el paso cuatro es la adaptación del peso, para lo cual se usa un algoritmo recursivo a la salida de los nodos trabajando hacia atrás en la primera capa oculta, ajustando el peso de la forma:

$$w_{ij}(t+1) = w_{ij}(t) + \mu \varepsilon_j x_i' \tag{B.16}$$

Donde $w_{ij}(t)$ es el peso desde el nodo oculto i o desde una entrada al nodo j en el tiempo t, x_i' es la salida del nodo i o una entrada, μ es el término de la ganancia y ε_j es el término de error para el nodo j, si el nodo j es un nodo de salida, entonces se tiene:

$$\varepsilon_j = y_i(1 - y_i)(d_j - y_j) \tag{B.17}$$

Si el nodo j es un nodo oculto interno, se tiene la siguiente expresión:

$$\varepsilon_j = x_j'(1 - x_j') \sum_k \varepsilon_k w_{jk} \tag{B.18}$$

Donde k indica todos los nodos en las capas superiores del nodo j, adaptando los umbrales de los nodos internos en una forma similar al suponer que tienen

pesos conectados o enlazados desde las entradas imaginarias teniendo el valor de uno. La convergencia es algunas veces más rápida y los cambios de peso son más suaves si el término del momento es sumado de la forma:

$$w_{ij}(t+1) = w_{ij}(t) + \mu\varepsilon_j x'_i + \gamma(w_{ij}(t) - w_{ij}(t-1)) \tag{B.19}$$

Donde $0 \leq \gamma \leq 1$, repitiendo los pasos del dos al cuatro los pesos y los umbrales convergen, las redes neuronales típicas proveen un alto grado de robustez y tolerancia a fallas, sin embargo una deficiencia de los algoritmos backpropagation es que requieren de un número elevado de datos de entrenamiento para su convergencia.

Apéndice C

Artículos y ponencias en Congresos

Simulation of a model of the basilar membrane in two dimensions for Spanish vowels

160th Acoustical Society American Meeting
Cancun, México.
15-16 Noviembre 2010.

The inner ear have the cochlea as the principal element, this is a biological element in the form of a snail, within which the mechanical energy is converted into electrical energy, this process is realized by the inner ear cells on the basilar membrane. This membrane response is different for different frequencies of excitation; the result of this process is the human audition. This paper shows a simulation of a model in two dimensions of the basilar membrane and its characteristic response when excited by the tow first formants of Spanish vowels; these formants are obtained by an analysis by mixtures of Gaussians.

Modelado de la membrana basilar como un sistema de osciladores armónicos

IV Congreso Nacional de Ingeniería en Comunicaciones y Electrónica
Ciudad de México, México.
24-26 Noviembre 2010.

El presente trabajo muestra la simulación de la respuesta del modelo de la membrana basilar como un oscilador armónico propuesto por Lesser y Berkley en 1972. El objetivo es determinar la relación que existe entre la frecuencia de excitación de un estímulo al sistema auditivo y la distancia a la cual la membrana basilar responde. Los resultados obtenidos permiten observar la relación que existe entre los valores de la masa y la constante de elasticidad para cada oscilador propuesto, con los valores de la frecuencia de excitación y la distancia a la cual responde la membrana basilar.

Modelado del Oído Interno aplicado al Análisis de Señales de Voz

19 Congreso Internacional Mexicano de Acústica
Ciudad de México, México.
5-7 Diciembre 2012.

El presente trabajo muestra la simulación de la respuesta del modelo de la membrana basilar como un oscilador armónico propuesto por Lesser y Berkeley en 1972. El objetivo es determinar la relación que existe entre la frecuencia de excitación de un estímulo al sistema auditivo y la distancia a la cual la membrana basilar responde, logrando con esto tener valores que sean utilizados como parámetros de análisis de señales de voz empleando un modelo de Wavelets Gaussianas. Se utilizan señales de prueba sinusoidales en intervalos de frecuencias de octavas desde 50 Hz hasta 2 KHz. Los resultados obtenidos permiten observar la relación que existe entre los valores de la masa y la constante de elasticidad para cada oscilador propuesto, con los valores de la frecuencia de excitación y la distancia a la cual responde la membrana basilar.

Computational model of the cochlea using resonance analysis

Revista Mexicana de Ingeniería Biomédica
Diciembre 2012, Volumen XXXIII, Número 2, páginas 77-86.

This paper presents the development of a computational model of the cochlea using a new solution by resonance analysis to the models of fluid mechanics in the cochlea and the basilar membrane as a system of forced harmonic oscillators proposed by Lesser and Berkeley. The computational model of resonance analysis is successfully compared with the method of numerical integration developed by Peterson and Bogert, the method of Green function proposed by Allen, the method of finite difference described by Neely and the measurements obtained in the experiments of Bksy, getting the same results with the new solution developed. Its contribution regarding the different solutions already found in the literature is to obtain a frequency-distance function to identify the maximum amplitude of displacement of each section along the basilar membrane for each specific excitation frequency in the hearing system. The model developed presents the advantage over the previous solutions, that the function obtained depends only of the physical characteristics of mass per unit area, damping coefficient and stiffness per unit area along the basilar membrane, and is the first time that the resonance analysis is used to obtain a methodology consistent with the place theory of hearing of Békésy.

Modelos del sistema auditivo aplicados a los sistemas de reconocimiento de voz

Expocacústica 2013.
Ciudad de México, México.
28 Enero - 1 Febrero 2013.

En los sistemas de reconocimiento de voz es posible usar la respuesta del modelo de la membrana basilar como un oscilador armónico para la parametrización eficiente de la señal de voz. Lo anterior se fundamenta en la determinación de la relación que existe entre la frecuencia de excitación de un estímulo al sistema auditivo y la distancia a la cual la membrana basilar responde, siendo esta metodología concordante con la teoría los puntos de audición de von Békésy. Los resultados obtenidos permiten observar la relación que existe entre los valores de la masa, factor de amortiguamiento y constante de elasticidad para cada oscilador propuesto, con los valores de la frecuencia de excitación y la distancia a la cual responde la membrana basilar. Siendo esta una metodología alternativa al cepstrum, el análisis predictivo lineal y los análisis perceptuales con base en la escala de Mel y la escala de Bark.